内 容 提 要

本书是一部数学专题史，以概率论思想方法的发展为主线，简要地阐述了古典、近代及现代概率论的发展历程，也对概率论在中国的传播与发展给出了一个大致的脉络。在介绍了许多对概率论的产生与发展做出过重大贡献的数学家的同时，对他们的工作予以恰当的分析与评价，从哲学、文化的视野探讨了概率论产生、发展的机理。内容有趣，富于哲理，给人以启迪。本书可供大、中学师生教学参考、课外阅读，也可供数学史、文化史爱好者阅读。

作 者 简 介

朱春浩，男，湖北广水人，毕业于哈尔滨工程大学理学院应用数学专业，硕士，副教授，中国数学会及数学史分会会员，湖北省数学会高职高专数学研究会常务理事、副理事长，武汉船舶职业技术学院学术委员会委员， 院学报编辑委员会委员，主要从事概率论与数理统计及发展史的研究。在《数学杂志》、《经济数学》、《统计与决策》、《高等数学研究》、《高等数学通报》等数学专业刊物上发表论文 20 余篇，主编、主审高等数学、经济数学、工程数学、电路数学等教材多部，主持了湖北省人文社会科学规划课题“数学、数学科学与人文社会科学关系研究”(课题编号：2006y346)、湖北省高等学校省级教学研究课题“高等职业教育数学教学整体优化模式的研究”（课题编号：20060407）、中国职业技术教育学会职高委教研会重点课题“新世纪高等职业教育工科类数学课程体系与教学内容的改革研究”等多项省部级课题的研究。

前　言

任何企图将一种科目和它的历史割裂开来，我确信，没有哪一种科目比数学的损失更大。

——格雷舍（英）

如果我们想要预见数学的未来,适当的途径是研究这门科学的历史和现状。

——庞加莱(法)

介绍科学成就的书籍应该不仅叙述人类思维有限的结果，而且要把读者引到研究工作过程的本身中去，告诉读者怎样克服困难，怎样寻求正确的方法。

——高尔基（苏）

史学的研究历来受到思想家们的重视，伟大的哲学家弗朗西斯·培根曾经说过“读史使人明智”，“温故而知新”更是每一个中国人从小耳熟能详的至理名言。因此，史学研究的重要性是不容置疑的。

概率论是研究随机现象的规律性的数学理论，产生于 17 世纪中叶。它不仅有自己独立的研究问题，还在现实世界中有着意想不到的应用。20 世纪以来，概率论逐渐渗入到自然科学、社会科学以及人们的日常生活等几乎无所不在的领域中。无论在研究领域，还是教育领域，它愈来愈成为一门当今最重要的学科之一。概率论历史的研究也日益引起科学史学家们的重视。关于概率论的历史，几乎所有教本都是一带而过，很少有详尽的论述，这是应该弥补的不足。任何一门学科的创造、形成以及发展都有其连续性，割断历史、忽略历史都会给教学和科研造成不应有的损失。

本书是基于数学哲学的研究，进行一种尝试：把数学是一种文化的观点融合到数学哲学问题的思辨中，并由此开拓出数学文化研究的新领域。也就是把数学科学作为人类文化活动及其成果，力图从内史与外史结合的角度进行研究。在充分借鉴、吸纳国内外有关学术理论成就的基础上形成数学文化的思想、观点和方法。

本来，对数学的哲学思考应该是关于数学与人类整体思想、文化，数学与宇宙和人类社会的关系的全范围和多角度的思考，但以往的数学哲学研究由于过多地纠缠于数学内部体系中的各种技术性、知识性和方法论的细节方面，而将原本意义重大、生动丰富且范围广泛的数学与社会、人文科学的关系领域忽略了！在我国很长一段时间里，对数学本质的认识，偏重于数学基础、数学与生产实际及自然科学的关系，忽略了它是人类文化的重要部分。

在我们进行数学哲学思考的时候，特别是当我们进行某些更深刻的对数学本质、意义的理解时，经常会被一些想法所困惑：我们在以往的数学哲学或相关的研究中，是否有一个重大的偏差，就是忽略或回避了数学的文化性。我们是否过分地强调了数学纯粹的科学性而把视野局限在其实是更广阔的数学文化研究领域的一角。受这些思想的驱使，不再把数学仅仅放在哲学的层面上加以理解，而是把数学看成人类所创造的全部文化的一个有机

部分来看待，并把探讨数学与其他文化的联系作为一项重要的课题加以研究。

过去的数学史，它更多的是告诉我们发生什么，而不是告诉我们怎样发生。本书试图对数学发展有着重要影响的哲学、文化观念进行数学文化史意义上的论述和探讨。这一研究方向将会随着社会发展和学校文化教育的日趋完善而日益显示出其重要性和迫切性。因为这是大学教育，是大学数学教育中实现文理沟通的非常有效的途径之一。是加强对数学发展有着重要影响的社会经济结构、哲学和文化观念的认识和探讨，有助于数学文化学的建设。意义是深远的。

本书在前人工作的基础上，通过对国内外大量概率论历史的研究资料进行分析，结合对原始文献的研究，试图以年代为线索，以代表人物为依托，比较系统地探讨概率论的发展状况，使我们对概率论的思想方法演变过程有个整体的了解，并以此为背景探讨我国概率论方面的最早的译著——《决疑数学》的一些历史问题。

本书从概率研究所使用的数学工具这一角度，尝试对概率论的历史进行划分，并且试图分析每个时期的特征。尽管这种划分有待商榷，界线也并不明显且有交叉。但从历史上看，作者仍将概率论经历划分三个主要发展阶段：古典概率论——组合概率论时期、近代概率论——分析概率论时期和现代概率论——测度概率论时期。

全书共分8章，各章内容如下：

第1章古典概率论——组合概率论：从17世纪中叶诞生至1812年，概率计算主要以代数方法为主，这一时期称为古典概率论。概率论起源于17世纪中叶人们对机会性游戏的数学规律的探讨。这个学科的发展与数学史上一些伟大的名字相联系，如费马、惠更斯、雅各布·伯努利、棣莫弗等。他们对这个专题的研究做出了重要的贡献。

第2章近代概率论——分析概率论：从1812年到20世纪初，主要以分析方法为主，如：特征函数、微分方程和差分方程等，这一时期称为分析概率论。

第3章现代概率论——测度概率论：1933年以后，主要以测度论来研究概率论，可以称为测度概率论，这时概率论已经实现了公理化。在公理化基础上，现代概率论取得了一系列理论突破和迅速发展。

第4章随机变量的分布产生的背景：随机现象统计规律的研究，可以转化为实变量的实值函数的研究。通过引进随机变量及其分布函数的概念，实现了这一转化。一般的统计类的书籍只是简要介绍各类分布的定义、性质和应用。而对于其产生的背景，即它是怎么来的，为什么要研究这类分布?没有一个详细的讲解。本章针对这种情形在查阅了大量中、外文资料后，对各类分布的产生背景进行了整理汇编，并用统一的符号进行叙述。

第5章极限理论： 概率论的认识论的价值只有通过极限定理才能被揭示，没有极限定理就不可能去理解概率论的基本概念的真正含义。概率论极限理论是概率论的主要分支之一，也是概率论的其他分支和数理统计的重要基础，极限理论是19世纪后期以来概率论研究的一个中心议题。

第6章概率论在中国的传播与发展：考察了原来一直有争议的《决疑数学》的原著的问题。通过详细的考证，发现《决疑数学》的唯一原著是在《大英百科全书》第八版中托马斯•伽罗威所作的“概率论”一文，而与原来人们所认为的《钱伯斯百科全书》（新版）中安德森的文章没有关系。鉴于《决疑数学》中所翻译的大多数概率术语并未保留下来，也没有开创一个概率论研究的传统。近代概率论在中国的传播和发展显然与《决疑数学》

并没有一个一脉相承的关系。因此得出结论：《决疑数学》对概率论这一学科在中国发展的影响是非常有限的。

第7章概率论发展中若干问题的思考：本章着重从以下几个方面进行理性的思考：概率论起源的哲学思考，“实践——认识——实践”是概率论历史发展的必然规律，随机性的认识论，从概率论的发展看东西方文化传统的差异，数学分析对概率论的渗透与推动，概率论与数理统计有着某种必然的联系，概率论在统计学发展中的角色研究。

第8章概率统计教学中渗透概率统计发展史：哲学与文化视角下概率统计课的育人功能，在概率统计教学中渗透数学史的作法与体会。

记住下列人物：卡尔达诺、费马、惠更斯、雅各布·伯努利、棣莫弗、拉普拉斯、高斯、泊松、切比雪夫、马尔可夫、李雅普诺夫、凯恩斯、伯恩斯坦、米泽斯、博雷尔、柯尔莫戈洛夫、莱维、辛钦、伊藤清、许宝騄等等对概率论的产生、发展所作的历史性贡献。

概率论历史的研究是一个内容丰富且比较复杂的研究领域，而概率论与数理统计的发展又有着千丝万缕必然的联系。更为详实的历史描述和全面的专题分析还有待将来去完成：

（1）概率论发展的文化、哲学方面的深入研究。

（2）数理统计史的全面而深入的研究。

（3）统计推断思想的历史研究。

（4）概率论与数理统计学的课程重建。

本书在写作过程中参考了大量的原始文献，也采用了一些同仁的研究结论， 主要分为三类：

（1）原始文献。

（2）有关概率论史研究的文献。

（3）其他方面的参考文献。

在此谨向这些文献的作者或译者表示真诚的感谢！

最后我要感谢我的妻子王建勤女士和女儿朱思妤。多年来，她们总是在我需要的时候无私地给予我精神上的支持。如果没有她们的理解、关心与照顾，我不可能安心地从事研究工作。我愿将此书奉献给她们！

本书是第一次比较系统地研究概率论的发展史，由于学历所限，书中还有不少疏漏甚至错误，敬请学界前辈、同仁和读者教正！如果您有什么意见或建议，可以通过电子信箱zhuchunhao@sohu.com 与我联系。

目 录

第 1 章　古典概率论——组合概率论

从 17 世纪中叶概率论诞生至 1812 年，概率计算主要以代数组合方法为主，这一时期称为古典概率论——组合概率论时期。

1.1 排列与组合

历史上排列和组合问题出现得很早。中国古代《易经》中的八卦符号“—”、“--”和“-”的可重复排列，共有 2^3 种。公元 1 世纪古希腊史学家普鲁塔克(Plutarch)曾告诉我们，公元前 3 世纪的希腊哲学家克里西普(Chrysippus，约公元前 280–前 207)曾发现 10 个公理的不同排列数超过 1 000 000，而公元前 2 世纪的希腊天文学家伊巴谷则错误地给出该排列数为 101 049 或 310 925。

中世纪意大利数学家博伊修斯(A.M.S.Boethius，408~524)曾给出 n 件物品中一次取 2 件的组合数公式，用我们今天的符号表示就是

$$C_n^2=\frac{n(n-1)}{2}.$$

公元 9 世纪，印度史学家摩诃毗罗给出一次从 n 件物品中取 r 件的组合数 C_n^r 的乘法法则。12 世纪，婆什迦罗在他的《丽罗娃底》中给出一次从 n 件物品中取 r 件的排列（可重复或不重复）数和组合（不重复）数 C_n^r 的算法，还给出总个数为 n 的 m 种物品（其中第 i 种物品的个数为 r_i ）的排列数

$$\frac{n!}{r_1!r_2!\cdots r_m!}(r_1+r_2+\cdots+r_m=n).$$

婆什加罗还知道下面的组合数性质

$$C_n^r=C_n^{n-r}.$$

在犹太数学文献中，有关排列与组合的内容出现得很早。在不迟于公元 8 世纪写成的《创造之书》(Sefer Yetsirah，作者不详)中，作者给出了 22 个希伯莱字母的全排列，称“两个字母可构成两个单词，三个字母可构成 6 个单词，4 个字母可构成 24 个单词，5 个字母可构成 120 个单词，6 个字母可构成 720 个单词，7 个字母可构成 5 040 个单词，依次计算下去……”

公元 946 年，多诺罗在注释《创造之书》时证明了 n 个字母的全排列数 $n!$ 。

12 世纪，犹太数学家伊本·艾斯拉研究过在已知行星中一次取两个、三个或更多个（会合）的组合数问题。艾斯拉知道，从 7 件中一次取 2 件的组合数与一次取 5 件的组合数相等，类似地，一次取 3 件和一次取 4 件的组合数、一次取 6 件和一次取 1 件的组合数也分别相等。他可能已经知道组合数一般公式和性质

$$C_n^r=C_n^{n-r}.$$

12世纪末、13世纪初，阿拉伯数学家伊本·穆尼从语言学里得出排列与组合的特殊公式(易于一般化)。13世纪末，阿拉伯数学家伊本·阿尔巴拿在《算术运算概论》及其自注中给出并证明了n个元素的全排列数P_n^n，一次从n个元素中取r个的排列数P_n^r和组合数C_n^r的公式，其中有

$$C_n^r = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}.$$

以及

$$C_n^r = \frac{n-r+1}{r}C_n^{r-1}.$$

书中一般公式的论证和运用表明，在13世纪的伊斯兰世界，人们已将组合公式当作工具，用于算术、代数、几何或天文学中。

14世纪，法国犹太数学家本·吉尔森在写于1321年的《数之书》中深入研究了组合问题，给出并证明了n件物品的全排列数以及一次取出r件的排列和组合数。法国数学家奥雷姆给出从6件物品中一次取出1件、2件、3件、4件和5件的组合数之和，但奥雷姆在计算一般组合数C_n^r时却出了错。

文艺复兴时期的欧洲，许多数学著作中出现了组合问题和法则。如帕乔里（Luca Pacioli，1445~1509）的《算术、几何、比例和比例性集成》(1494)、卡尔达诺《探微》(1550)和《比例新论》(1570)、塔塔格里亚《数量通论》(1556)、布丢《算法》(1559)和巴克莱的《记忆算术》(1567)等。帕西沃里在其书中给出了如何求坐在一桌的任何多个人的全排列数。英国数学家巴克莱在其书中给出求n件物品一次取出r件的组合数的特例。卡尔达诺获得了公式

$$C_n^r = \frac{n-r+1}{r}C_n^{r-1}.$$

而法国数学家布丢则不仅讨论了四骰子的组合问题，而且还研究了带有若干可转到圆柱的组合问题。

在同一时期，阿拉伯数学家柯多维罗在自己的著作中对排列组合也有论述，并应用了若干一般公式。

17世纪，组合研究达到了一个全新的水平。法国数学家德·贝西在《组合简法》、斯特罗德在《组合、选择、排列与量的合成》中都对该课题做了研究。在1634年出版的著作中，法国数学家埃里岗给出了组合数的一般公式

$$C_n^r = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}.$$

帕斯卡在《论算术三角形》中将他的算术三角形应用于组合理论，其主要结果是：

(1)在底边有n个单元的算术三角形中，第$r(1 \le r \le n)$行的所有单元之和为n件物品中取r件的组合数，此即

$$C_{n-1}^0 + C_n^1 + C_{n+1}^2 + \cdots + C_{n-1}^{n-r} = C_n^r.$$

(2) n件物品中一次取1个、2个、……、n件的组合数之和为

$$C_n^1 + C_n^2 + C_n^3 + \cdots + C_n^n = 2^n - 1.$$

(3) n 件物品中一次取 r 件的组合数为算术三角形第 $n+1$ 条底边上位于第 $r+1$ 行的单元，即

$$C_n^r = \frac{(r+1)(r+2)(r+3)\cdots n}{(n-r)!}.$$

或

$$C_n^r = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}.$$

帕斯卡和英国数学家沃利斯最早使用了“组合(combination)”这一术语。

同一时期的法国数学家范・舒腾(F.Van Schooten，1615~1660)也探讨过组合数问题。他首先取 4 个字母 a,b,c,d，列出其中 1 个、2 个、3 个和 4 个的所有可能的组合如下：

a.
b.　*ab.*
c.　*ac.*　*bc.*　*abc.*
d.　*ad.*　*bd.*　*abd.*　*cd.*　*acd.*　*bcd.*　*abcd.*

共有 15 种组合。

由此，范・舒腾得到，如果 a,b,c,d 是一正整数的四个不同因素，则包括 1 在内它有 16 个因素。然后，范・舒腾取 5 个字母，共有 31 种组合，最后他得出一般结论：取 n 个字母中 1，2，3，…，n 个的所有可能的组合，总数为

$$2^n - 1.$$

设 a,b,c,d 是四个不同素因数，则易知 $abcd$，a^3bc，a^3b^3，a^7b 和 a^{15} 都有 16 个因数。由此，范・舒腾提出如下问题：在含有 16 个因数的所有正整数中，哪一个是最小的?

范・舒腾的正确答案是 $2^3 \cdot 3 \cdot 5 = 120$。

类似地，a^2bcd，a^3b^2c，a^5bc，a^5b^3，a^7b^2，$a^{11}b$ 和 a^{23} 都有 24 个因数，$2^3 \cdot 3^2 \cdot 5 = 360$ 是这些正整数中最小的一个。

德国数学家莱布尼兹(G.W.Leibniz，1646~1716)在 1666 年发表的论文《论组合的艺术》中，对组合数和全排列做了研究。他构造了一个类似于帕斯卡三角的数表，并利用它来求组合数。莱布尼兹将一组物品中一次取 2 个的组合数记为 con2natio，一次取 3 个的组合数记为 con3natio，一次取 4 个的组合数记为 con4natio，等等，但莱布尼兹并没有像帕斯卡那样明确给出组合数的一般公式。莱布尼兹证明了如下定理：若 n 是素数，则 n 个元素中一次取 r 个的组合数能被 n 整除。莱布尼兹似乎并不知道帕斯卡的有关工作，就组合理论而言，其论文水平也不及帕斯卡的《论算术三角形》高。

英国数学家沃利斯在其《代数学》(1685 年)附录“论组合、排列与除不尽部分”中专论排列与组合问题。沃利斯也利用等价于帕斯卡三角的数表来求组合数。他不仅计算不重复字母的全排列(如 Roma 一词中的四个字母的全排列数为 $1\times2\times3\times4=24$)，而且还计算有重复字母的全排列(如 Messes 一词中的 6 个字母的全排列数为

$$(1\times2\times3\times4\times5\times6)/(1\times2\times2\times3)=60.$$

17 世纪第一个有关组合的著名例子值得一提。1617 年，一位名叫普泰努斯(E.Puteanus)的作者在荷兰安特卫普出版了一部名为 Erycci Puteani Pietatis Thaumata in Bernardi Bauhusii

e Societate Jesu Proteum Paethenium 的书，书中含有歌颂圣母玛丽亚的诗句

Tot tibi sunt dotes，Virgo，quot sidera coelo

的 1022 种不同排列(诗句仍然有意义，但不能表达对圣母的赞颂的排列除外)，占了整整 48 页，如以 Tot tibi 开头的排列共有 54 种，以 Tot sunt 开头的共有 25 种等。有趣的是，1022 恰恰是古希腊托勒密星表中的恒星数。沃利斯以及后来的雅各布・伯努利对此都有讨论。

雅各布・伯努利在《猜度术》第二部分中论排列和组合，证明了一组元素的不重复或有重复的全排列以及 n 个元素中一次取 r 个的组合数与排列数的公式。这些内容，正如伯努利自己承认的那样，已经为范・舒腾、莱布尼兹、沃利斯等人研究过，并不完全是新的。伯努利首次创用“排列”(permutation)这一术语。他还证明了指数为正整数的二项式定理，并推导了自然数幂和公式。

1.2 概率论的孕育

现在常常认为概率论以 1654 年帕斯卡与费马之间的通信中对“点数问题”的解决为标志而突然产生，这种说法未免太简单。实际上，考古证明，早在古代就有了骰子，人们通过它来预测未来，占卜命运，赌博游戏。那为什么直到 17 世纪概率论才诞生呢?这是因为当时的社会背景有其特殊之处，有新的问题、动机来推动新理论的发展。那些具有概率性质的最初问题，起源于人类生活的各种领域，后来逐渐具体化成概率论的概念和方法。

17 世纪荷兰、西班牙、法国、英国、德国首先出现了各种参考手册，上面记载着教区居民结婚、参加洗礼、举行葬礼的登记数。这是在瘟疫流行的时候引进的方式，最早的可追溯到 1517 年。后来还增加了出生、死亡人口的性别以及死亡原因等。基于这些统计资料出现了一些概念，比如在某一阶段的可能性，能活到某一年龄的机会等等。毫无疑问，哲学发展影响了概率论基本概念的形成。因此，在各种历史时期里，一定程度上进行着收集、分析统计数据的活动。然而，直到出现资本主义，系统而足够广泛的统计研究才开始，那时贸易和货币交易，尤其是与保险有关的业务正迅速发展，而且各种新机构相继建立。

可以说，统计是推动概率论早期发展的一个基本因素。资本主义关系的增长不断地提出新的统计问题，而资本主义时期是从 16 世纪才开始的。

14 世纪荷兰、意大利率先建立了海运保险公司。海运货物保险是继陆地、湖泊、河流运输保险之后而出现的。这些公司通过计算各种风险，收取相应的保险金。海运的保险费是货物价值的 12%~15%，而陆地的费用是货物的 6%~8%。从 16 世纪开始，许多国家也出现了海运保险公司，17 世纪其他保险形式也相继诞生。保险公司收集的数据成为概率论发展初期所利用的原始材料。

文艺复兴时期自然科学迅猛发展，观测和实验的重要性也日益增加，处理观测数据的方法，特别是估计观测中出现的误差，这些问题就变得重要了。这也刺激了概率论的发展。

偶然性和必然性之间的相互关系，规律和因果关系等问题都是古代研究的对象，长期以来列在哲学家的研究议题之中。因此，毫无疑问，17 世纪所积累的哲学思想影响了概率论的早期发展。

人口统计问题，保险公司的实际需要，处理各门科学的数据或观测结果，以及与机会游戏有关的抽象问题，它们与组合学，与偶然性、必然性的哲学观念有着密切关系，刺激

了概率论的诞生。

概率论起源于 17 世纪中叶，当时在误差、人口统计、人寿保险等范畴中，需要整理和研究大量的随机数据资料，这就孕育出一种专门研究大量随机现象的规律性的数学，但当时刺激数学家们首先思考概率论的问题，却是来自赌博者的问题。

随着赌博的盛行，人们开始研究赌博中的规律。赌博问题研究的重要意义并不在于解决问题本身，而在于通过研究使人们对概率的性质有了一个较为深刻的认识，并最终导致概率概念的形成。当时，人们研究问题的习惯方式是举一些有实际意义的例子。在赌博的问题中，对后世影响最大的当数“赌本分配”(division problem)问题。所谓的赌本分配问题是这样的：两个赌徒 A 、 B 事先约定进行若干局的公平赌博(公平赌博意味着双方获胜的机会都是 1/2，双方所出的赌资也都一样)，直到其中一人赢了某一事先规定好了的局数为止，比方说 s 局。现在由于某些偶然因素导致赌博无法进行，此时 A 赢了 s_1 局，B 赢了 s_2 局($s_1 < s$ ，$s_2 < s$)，问应如何分配赌本才算公平?赌本分配问题也称“点数问题” (problem of point)，这是因为我们可以给每局的获胜者赋予 1 点，用点数的计算来代替局数的计算。

赌本分配问题最早见于意大利数学家帕乔里 1494 年所出版的一本书中，书名为《算术、几何、比与比例集成》。书中帕乔里就 $s_1 = 5$ ，$s_2 = 2$ ，$s = 6$ 的情况进行了讨论。他认为，赌博结束最少要进行 s 局，最多进行 $2s-1$ 局，因此应按 $\frac{s_1}{2s-1}:\frac{s_2}{2s-1}$ 或 $s_1:s_2$ 的比例分配赌金。帕乔里的推理没有包含任何的概率的理论和组合原理，因此对后世解决此问题没有任何意义。帕乔里之后，在解决这个问题上有进展的就是意大利数学家卡尔达诺。

卡尔达诺(Cardano Jerome，1501~1576)，1501 年 9 月 24 日生于意大利帕维亚(Pavia)，是一个法官和一个寡妇的私生子。自幼体弱多病，备受歧视和虐待，性格冷漠倔强。父亲法齐奥(Fazio Cardano，1444~1524)博闻饱学，在米兰讲授过法学和医学，曾与意大利文艺复兴时期的著名画家、科学家达·芬奇(Leonardo da Vinci)为友。受父亲的鼓励，卡尔达诺开始学习古典文学、数学和星占学。1520 年在帕维亚上大学时又学习医学。后转学于帕多瓦，1526 年毕业，取得医学博士学位，继而在帕多瓦附近的一个小镇萨科隆戈(Saccolongo)行医近 6 年。1531 年与 L.班达雷妮(Bandareni)结婚，生有二子一女。婚后不久，卡尔达诺因收入微薄，难以支撑不断扩大的家庭，被迫搬到米兰居住以谋公职。但由于他是私生子，米兰医学协会认为这是出身卑贱，拒绝他加入该协会。卡尔达诺只好独自开业行诊，生活十分拮据。1534 年由父亲的一个贵族朋友举荐，卡尔达诺成为米兰专科学校的一名数学教师，在那里讲授几何学。同时任贫民院的医生，生活略有好转。他除了教学和诊病外，还潜心医学研究。自 1536 年起在威尼斯等地出版了几部专著，阐述一些理论问题，总结行医经验，还揭露过医学界的某些劣行。由于他的医术高超，逐渐在米兰取得声望。1539 年米兰医学协会重新决定接纳他为该协会正式会员。同年卡尔达诺转到米兰的医学院任教，不久成为该院的负责人。1543 年又到帕维亚大学任医学教授。几年之内，成为闻名全欧的医生。1552 年还专程到英国爱丁堡为大主教 J.哈密顿(Hamilton)及其他达官显贵治病。1560 年，卡尔达诺宠爱的大儿子因犯“毒死妻子罪”被处决，对他的精神打击很大。当时，卡尔达诺的小儿子也生活放荡，桀骜不驯。为摆脱烦恼，卡尔达诺谋到波伦亚(Bologna)大学医学教授的职位，1562 年正式赴任。卡尔达诺的坎坷经历使他的性格颇为奇特，因而常常被描述为科学史上的怪人。他在数学、哲学、物理学和医学中都有一定成就，同时也一直醉心于占星术和赌博的研究。1570 年因给耶稣算命(说耶稣的一生都是受天上星宿的支配)

而受到宗教法庭监禁，被起诉为异教徒(另一说是因为债务问题被捕入狱，还有的说二者兼而有之)。几个月后，宣誓放弃异端邪说获释出狱，但失去了教学职位和学术出版权。1571年移居罗马，另谋生计。后因星占学研究得到教皇皮乌斯五世的赏识，付给他终身年薪，留在皇宫供职。在生命的最后一年(1576)，卡尔达诺写下了自传体著作《我的生平》(De propria vita)。卡尔达诺长期醉心于游戏、赌博、掷骰、弈棋、打牌。他早在1539年的著作中就论及赌金分配问题，另外又写成经验之谈式的专著《游戏机遇的学说》(The Book of Games of Chance)。这部著作直到1663年才收入在莱顿出版的10卷本卡尔达诺《全集》(Opera omnia)中第一次发表。书中给出一些概率论的基本概念和定理，得到所谓“幂定理”(某事件重复n次发生的概率)和大数定律。但这些理论发表得较晚，对后世影响不大。

卡尔达诺于1576年9月21日在罗马逝世，享年75岁。他的一生既有着辉煌成就又历尽沧桑，饱受世人责骂和侮辱，可以说集毁誉于一生。

卡尔达诺在数学方面造诣很深，特别在概率统计方面，作出了开拓性的贡献。1539年，他向著名数学家塔尔塔利亚(Tartaglia Niccolo，1500~1557)求教三次方程式的解法，并立誓予以保密。但他并没有遵守诺言，于1545年出版《代数规则大艺术》（Art Magna，简称《大术》）一书，将三次方程式公之于世。后来，三次方程式求根公式被命名为“卡尔达诺公式”，方程的解法被称为“卡尔达诺法”，而塔尔塔利亚在这方面反而湮没无闻。为了此事，卡尔达诺长期遭到学术界和舆论界的批评，被称为“不道德的人”、“剽窃者”，甚至是“臭名昭著的骗子”。不过，400多年过去了，人们开始对卡尔达诺有了重新评价，卡尔达诺在数学上并非平庸之辈，他的《大术》包含了许多他本人的独特创见，例如，他最早认识到负数和虚数， 认真地加以讨论并给出了表示虚数的符号和运算法则；他对代数方程论（包括三次方程）的研究有着重要的推进，并研究了四次方程的特例。

卡尔达诺在1539年出版的一本书中开始着手解决赌本分配问题。他首先意识到，公平地分配赌本原则只与双方为获胜所需赢得局数$a=s-s_1$，$b=s-s_2$有关。为此他引进了一个新的赌博游戏，双方从零开始，每局双方获胜的概率相同。如果A在B赢得b局之前赢得a局，则A获胜，反之亦然。他问双方如何下注赌博才算公平的?进而他认为新问题中双方的赌注比例就应是赌本分配问题中双方分配赌注的比例。他的结论是这个比例式

$$b(b+1):a(a+1)$$

虽然卡尔达诺的推理已包含了一些概率的踪影，但还是相当模糊的，并且他导出的分配原则也不正确。不过他引进的新赌博游戏却为后人正确求解指明了正确的方向。

卡尔达诺的一本与概率论发展有关的著作出版于他死后的1663年，书名为《游戏机遇的学说》。

他的《游戏机遇的学说》一书最早探讨了有关机遇的数学理论。这一著作大约完成于1564年，是他根据多年的赌博经验和数学理论写成的，主要内容是作者关于赌博中的道德情操、赌博实践、赌博理论的总结，费洛伦斯·戴维把它叫做是“理论数学家和赌徒的结合体”。

1. 内容与体例

全书共由32个较短的章节组成，分述如下：

前5章是入门介绍，讲述游戏的种类，玩游戏的条件，参与游戏的人员与时机，游戏可能产生的结果以及写作此书的缘由，从中可以看出卡尔达诺写作态度是比较认真的。他首先从纸草，铭文碑刻，一直列举到蜡版与树皮文献，简单阐述了赌博历史。接着在第2

章中申明赌博游戏的利害，如可带来忧虑、怒气、引起争吵等。他从医生的角度提出忠告：游戏只是一种娱乐，不会给失意者带来任何安慰。在第 3 章中卡尔达诺更明确指出，有名望的人、老年人和教士等人不宜参与赌博游戏，它只是孩子、年轻人和士兵等人的活动，而且是短暂的、适当的活动。特别指明不能与职业赌徒进行游戏，那将是不光彩和危险的。在第 5 章卡尔达诺说明了写作此书的两个理由：第一，大量的人沉溺于赌博游戏是一种自然的罪恶与不幸(natural evil)，它需要一位医生像对待不治之症一样来讨论它，确切地指出在不同条件下，获胜的可能性有多大，以明示众人；第二，这是哲学家的习惯，找出事物的两面性，讨论赌博并非为了赞扬它，而是指出其有利与不利，并将后者减少到最低限度。

第 6~10 章讲述游戏的基本原理。卡尔达诺认为所有赌博游戏最基本的原理是平等的条件。例如对双方人员，旁观者，赌金，赌场，赌具中的骰子，骰盒等诸多条件全部是平等的，任何偏颇都会导致结果的不公正。他以自己的经验揭露了一些做假手段，例如旁观者的暗示、鼓动，对手的假动作，骰盒的倾斜等，特别指出骰子本身掺假对赌徒是非常危险的，他主要介绍两种骰子：一种是常见的立方体 6 面骰，每个面依次标有 1~6 个点，相对两面的点数之和为 7；另一种是 4 个面的距骨骰，每个面代表一种点。后者是考古发现中最早的游戏工具之一。

第 11~15 章包含了卡尔达诺关于掷骰理论的主要结果，分别讨论了同时掷两个骰和同时掷三个骰时各种可能情形数的计算和得点数的计算，以及多次同时掷两个或三个骰时可能情形数和点数计算。他针对当时通行的游戏规则，具体计算了求其一点数出现的可能次数，指出应依据有利情形出现的可能性大小下赌注，这样可以避免赢少输多的局面，并具体举例分析了通常容易产生错觉的赌博方式。

第 16~20 章论述打牌游戏，详细介绍了在 16 世纪盛行于欧洲的一种纸牌游戏 Primero 的玩法，包括 5 种“叫牌”的方法和三种成为“新牌”的情形，给出在平等条件下各种获胜可能性的大小，建立了平均值法则作为判别标准。他在第 17 章中揭露了牌游戏赌博的种种欺骗做假行为，第 20 章又以自己的亲身经历讲述了如何被别人欺骗，又如何靠智慧才能赢得最后的胜利的一段往事。

此后，两个较短的章节分析了掷骰游戏中的胆怯心理和游戏种类的细分方法。从第 23 章开始卡尔达诺又回到牌游戏的论述，他指出牌游戏与骰游戏的不同，前者带有隐蔽性，而后者是公开的。卡尔达诺早在大学毕业时就注意并喜爱上了打牌，因为这种游戏需要智慧和机敏，要根据已掌握的牌推测进一步的行动取舍，但他同时指出牌游戏中的做假现象太多，仅辨认牌背面来做假就“有一千种方法”，这不同于单纯靠运气的骰游戏。虽然如此，他还是客观地描述了公平打牌的各种技巧。

第 30 章是赌博游戏的回顾，卡尔达诺引经据典，从古罗马诗人马提雅尔(Martial，约 38/41~约 104)的记载，到希腊正教帕撒罗尼迦都主教优斯塔修斯(Eustathius，12 世纪初~约 1194)的描述，给出希腊传说人物帕拉墨得斯(Palamedes)在特洛伊 10 年围城战役时发明掷骰游戏之说，其目的是为了给士兵单调的生活增加娱乐。古代游戏只用一个骰子，后来才有了多个骰一起掷的玩法，这增加了判断的难度，却为探索其规律性提供了模型。

第 31 章专论距骨游戏，距骨一般指动物裸关节处的骨头，有 6 个面，其中两个面很小，所以距骨看上去只有 4 个面，其中一个微隆，其相对的面稍凹，四个面分别代表的数目是 1 和 6，3 与 4(两两位于相对的面)，卡尔达诺计算了连续 4 次掷距骨出现的各种困难

点数。继而，在最后一章(第 32 章)中作出总结，由于距骨不对称，因而计算结果与实际情况会有差异，卡尔达诺的这一工作只有理论价值。

其中与掷骰子有关的是 9~15 及 31~32 章。《游戏机遇的学说》被许多数学史家称为“赌徒守则”，可从以上介绍来看更像一本严肃探索赌博游戏理论的学术著作。它并不鼓动人们如何投机取巧，从中渔利，也没有介绍任何一种做假的方法，而是从多方面分析了游戏的利弊，提出了一系列公平游戏的法则，颇有君子风度。全书以古罗马政治家、作家和学者西塞罗(Cicero，公元前 106~前 43 年)在《论演说术》中的话作为结束语：“玩距骨或骰游戏只是在人们因天气原因不能干活时或劳作之余从事的一种游戏。”丝毫没有诱导人们投身赌博的意图。

2.主要贡献

(1)卡尔达诺首先申明骰子必须是“公正的”，这样作为立方体的 6 面骰在掷时任何一面向上的机会均等。

在书中，卡尔达诺谈到了掷距骨（踝骨）和骰子的机遇问题，通过理论上的论据提出了系统的概率计算。他说：踝骨有四面并有四个点子。但骰子有六面，抛掷六次，每个点子会出现一次。但因为有的点子出现不止一次，因而有的点子就出现不了。踝骨呈现扁平面，它的每一面位于背面之上。它不具备骰子的形式。骰子各面的总数之半总是出现相等次数，因而某个给定的点子抛掷三次的机会是均等的，因为总的变化是完全按六面顺序的。再者，在一次抛掷中，三个给定点子中之一将会出现。卡尔达诺由此推论，如果骰子不假，即如果每一面可以给予相等的权数，则我们可以计算机遇。书中明确指出，如果骰子是可靠地（honest），则六个面出现的机会相等，并且他还把“机会”（chance）定义为有利场合数与所有等可能的结果数之比。卡尔达诺的结论无疑是正确的。一颗六面完全相等的不假的骰子，每一面出现的机会，是可以通过数学抽象作出概率计算的，他还实际地计算了掷两颗或三颗骰子时在一切可能方法中有多少方法可以得到某一总点数。

(2)卡尔达诺还已知并使用了概率计算中的乘法原理，这是他书中最先进的结果。

该书还利用加法原则讨论了掷两枚或三枚骰子出现的可能性问题（表 1—1）。

表 1—1

有利场合数	两枚骰子	三枚骰子
包含 1 点	11	91
包含 2 点但不含 1 点	9	61
包含 3 点但不含 1、2 点	7	37
包含 4 点但不含 1、2、3 点	5	19
包含 5 点但不含 1、2、3、4 点	3	7
仅包含 6 点	1	1
总计	36	216

除了利用加法原则计算有利场合数外，卡尔达诺的另外一个领先于时代的结果是利用乘法原则发现了重复赌博中的赔率。作为一个职业赌徒，卡尔达诺当然更习惯用“赔率”(odds)一词来代替“机遇”(chance)或“概率”(实际上，无论是在古代还是在现代，人们在赌博时都更习惯用赔率代替概率)。赔率这个词对于喜欢参与足球博彩的人来说是再熟悉不

过了。赔率与概率的关系可以用一个简单的例子来说明：比方说，如果阿根廷夺取世界杯的赔率是 1：5 或 1 赔 5，那么阿根廷夺取世界杯的概率就是 1/(1+5)=1/6。

卡尔达诺考虑了如下情况：设一局赌博中所有可能结果数为t，有利结果数为r，则赔率为$r:(t-r)$。问题是重复地进行n局这样的赌博，赔率是多少(也就是n局全部获胜才算获胜的赔率)?经过反复摸索，卡尔达诺终于发现正确的答案是

$$r^n:(t^n-r^n),$$

令$p=r/t$，则上面问题的现代表述方法为

$$p^n:(1-p^n).$$

卡尔达诺除了上述对概率论的贡献外，他还在 1570 年利用递推公式

$$C_n^k=\frac{n-k+1}{k}C_n^{k-1},$$

导出组合公式

$$C_n^k=\frac{n(n-1)(n-2)\cdots(n-k+1)}{1\cdot 2\cdots k}.$$

但他并没在《游戏机遇的学说》中涉及到这个公式，说明他还没意识到组合理论对计算等可能情况下古典概率的重要性。

(3)利用平均值法则进行计算。对于 6 面骰，各面的点数之和为 21，被 6 除后得$3\frac{1}{2}$，即每一掷的平均得点数为$3\frac{1}{2}$。当三个骰子一起掷时，最大点数是 18，最小点数是 3，和为 21，则平均数是$10\frac{1}{2}$。

卡尔达诺重点分析了距骨游戏中的平均值计算。距骨骰的 4 个面分别是 1，6，3，4 点。1 被认为是不幸运而 6 则表示幸运。这种游戏常常是 4 个骰一起掷。其点数的组合只有 35 种情形，而排列则有 256 种。出现 1，3，4，6 情形称为维纳斯(Venus)，共有 24 种排列，出现的概率约为$\frac{1}{11}$。

4 个距骨的最大点数是 24，最小点数是 4，和为 28，则平均值为 14。这样，同时掷 4 个骰时，只要有多于一点出现(即至少两个一点)，这一掷就是不利的。因为其他两个骰无论是什么数，这一掷都不能超过平均数。人们掌握了平均数法则，再了解前述的可能情形计算，就可以在游戏赌博中掌握一定的主动权，从而选择有利的下注方式。

3.历史评价

卡尔达诺是将数学理论应用于赌博游戏研究的先驱之一。在他之前这种研究只有零星结果，例如掷三个骰子有 56 种组合方式发现于 10 世纪，有 216 种排列方式发现于 13 世纪。但卡尔达诺是第一个系统论述掷骰游戏的数学家。他用数学语言将所谓运气、机遇等经验公式化，创立掷距骨游戏中的各种概率大小。首次论述纸牌游戏中某些抽签的概率，还具体计算连续二次或三次掷两个或三个骰子的得点方法，并指出这种方法适用于任意次掷骰，是概率中“大数定律”的雏形，他不仅懂得如何实际参与赌博游戏获利，而且善于总结理论规律以指导实践，这在当时是罕见的。

卡尔达诺对掷骰问题所作的概率研究和计算，早于帕斯卡和费马建立概率论100多年，表明他是概率论的创始者。但是，他的这一著作在他生前并未发表，而是在他死后87年于里昂出版。这时书中的内容已为当时数学家所熟知(当时数学家已知道了加法公式、乘法公式、数学期望和一些组合理论，且赌本分配问题已由帕斯卡与费马于1654年解决)，因此这本书对后世影响不大。不过我们从书中可以看出，在卡尔达诺时代，人们对概率的认识是非常有限的。

伽利略(Galileo，1564~1642)在早期概率发展史上也占有一席之地。在1610年前后，为了回答佛罗伦萨宫廷贵族的赌博问题，他写了短文《关于骰子游戏的思想》，解决了以下问题：同时掷三个骰子，点数和为9的情形有6种(1，2，6)、（1，3，5）、（1，4，4）、（2，2，5）、（2，3，4）和（3，3，3）。点数和为10的情形也有6种(1，3，6)、(1，4，5)、(2，2，6)、(2，3，5)、(2，4，5)、(3，3，4)，那么出现点数和为9与10的机会应相同，而经验告知，出现10的机会比出现9的机会要多，原因何在?伽利略利用列举法得出了三个骰子的所有和数的不同，指出出现点数和为9和12有25种不同掷法，而出现点数和为10和11有27种，从而解释了为什么后者更容易出现的原因。伽利略也直觉地使用了等可能的思想，可见，已经产生了概率论的某些萌芽。

1.3 概率论的诞生

帕斯卡（Blaise Pascal，1623~1662）是法国数学家、物理学家、哲学家、散文家。1623年6月19日生于克莱蒙费朗；1662年8月19日卒于巴黎。

帕斯卡4岁丧母，其父是政府的官吏，博学多才，是一个业余数学家。由于帕斯卡从小体弱多病，其父不让他过早接触数学，以免思虑过度有损健康。帕斯卡12岁时，看到父亲阅读几何，便问几何学是什么，父亲为了不想让他知道得太多，就简单地告诉他几何是研究图形的，并且很快把数学书收藏起来，怕帕斯卡去翻阅，父亲对他接触数学的“禁令”，更激起了帕斯卡对数学的好奇心。于是帕斯卡就自行研究，当他把自己的发现：“任何三角形的三个内角和都是一百八十度”的结果告诉父亲时，父亲惊喜交集地流出了激动的眼泪，并改变了原来的想法，提早让帕斯卡学习《几何原本》等经典数学名著。帕斯卡贪婪地很快读完了《几何原本》。

1653年他写成了《论算术三角形》，经费马修订后于1665年出版，在这本书中建立起概率论的基本原理和有关组合论的某些定理。并与费马共同建立了概率论和组合论的基础，给出了关于概率论问题的系列解法。莱布尼茨后来读到帕斯卡这方面的研究成果时，深刻地意识到这门“新逻辑学”的重要性。

帕斯卡认为：“一个人的美德决不能从他特别的努力来测度，而应该从他每天的行为来测度”。他还说：“你要人们赞美你吗？那么你不要称赞你自己”。他认为：“数学是对精神的最高锻炼”。

帕斯卡在早年就表现出了超常的数学能力，是一位在科学史上富有传奇色彩的人物，18世纪的大数学家达朗贝尔(D’Alembert，1717~1783)赞誉他的成就是“阿基米德与牛顿两者工作的中间环节”。在数学史中他被称作“最伟大的天才”(Greatest Might-Have-Been)，他曾经对微积分、射影几何、概率论等数学分支做出了巨大的贡献。他拥有如此高的数学天赋和非常敏锐的直觉能力，他理应创造更多的发现。不幸的是，在他生命的大部分时间

里，他倍受敏感性神经痛和精神幻觉症的折磨。他于 1662 年去世时年仅 39 岁。

费马（Pierre de Fermat，1601~1665）是法国数学家。1601 年 8 月 20 日（另一说 17 日）生于图卢斯附近的波蒙特；1665 年 1 月 12 日卒于卡斯特尔。

费马出生于皮革商人家庭，他在家乡上完中学后，考入了图卢斯大学，1631 年获奥尔良大学民法学士学位，毕业后任律师，并担任过图卢斯议会议员。虽然数学只是他的业余爱好，但他对解析几何、微积分、数论、概率论都作出了杰出的贡献，被誉为“业余数学家之王”。费马在 1654 年写的一批信件中，他还同帕斯卡共同建立了概率论的一些基本概念。

费马性情谦抑，好静好癖。他对数学的许多研究成果都不愿发表（他的儿子在他去世后，才将其著作、信件、注记汇集成书出版）。这不但使他当时的成就无缘扬名于世，并在他的暮年也脱离了数学研究的主流，所以直到 18 世纪费马还不太知名。然而进入 19 世纪中叶，随着数论的兴起，数学家和数学史家对费马及其著作产生了浓厚的兴趣，争先发表研究费马的著作，其中尤以查尔斯·亨利（Cherles Henry）和保罗·坦纳（Paul Tannery）的 4 卷论文集最为全面，从中可以看出费马对数学和光学所作出的广泛而杰出的贡献。美国数学史家贝尔（Bell）说：“费马是一个第一流的数学家，一个无可指责的诚实人，一个历史上无与伦比的数论学家”。在数学中以他的名字命名的有：费马大定理、费马小定理、费马数、费马原理、费马螺线等。

尽管有卡尔达诺和伽利略等先驱者的一些非常重要的工作，而概率论历史学家大多赞同这样一个观点：对于数学中一个非常特别的问题的解法的探求成为数学化的概率科学产生的标志之一，这个问题被称作“点数问题”。所谓“点数问题”是指当游戏在完成前被终止时，怎样处理两名技能相当的游戏者的赌金分配问题，其依据是游戏者的得分数或是游戏终止时的点数。意大利的帕乔里早在 1494 年出版的《算术、几何、比与比例集成》（Summa de Arithmetica）一书中，就提到了赌博中常常遇到的“点数问题”，他是最早在数学著作中提到点数问题的作者。紧接着，卡尔达诺和他的对手塔尔塔利亚都讨论过这个问题。然而，所有这些人，对这一问题得出的结论都不正确。直到一百多年后，在 1654 年，一个名为德·梅雷（de Mere，1607~1684）的法国人把这个问题寄给了当时的数学天才帕斯卡，从此概率论历史上一个决定性的阶段才开始了。

德·梅雷是一位军人、语言学家、古典学者，同时也是一个有能力、有经验的赌徒，他经常玩骰子和纸牌。虽然他不是一个全职的数学家，但他经常从数学的角度提出和思考赌博中出现的一些有深度的问题，“点数问题” 就是其中之一。这一次，德·梅雷的问题的形式是：假设两个赌博者（德·梅雷和他的一个朋友保罗）每人出 30 个金币，两人各自选取一个点数，谁选择的点数首先被掷出 3 次，谁就赢得全部的赌注。在游戏进行了一会儿后，德·梅雷选择的点数“5”出现了 2 次，而保罗选择的点数“3”只出现了一次。这时候，德·梅雷由于一个紧急事情必须离开，游戏不得不停止。他们该如何分配赌桌上的 60 个金币的赌注呢？保罗认为，既然掷出他选择的点数的机会是德·梅雷的一半，那么他该拿到德·梅雷所得的一半，即他拿 20 个金币，德·梅雷拿 40 个金币。然而德·梅雷争执到：再掷一次骰子，对他来说最糟糕的事是他将失去他的优势，游戏是平局，每人都得到相等的 30 个金币；但如果掷出的是“5”，他就赢了，并可拿走全部的 60 个金币。在下一次掷骰子之前，他实际上已经拥有了 30 个金币，他还有 50%的机会赢得另外 30 个金币，所以，他应分得 45 个金币。

到底谁的分法对呢？当时可使两位费了不少脑筋，历史上古典概率正是由研究诸如此

类的赌博游戏中的问题引起的。现在我们一起来求解，显然，为确保能分出胜负，最多需要再赛两局，为简单计，用“+”表示“德・梅雷胜”，用“-”表示“保罗胜”，于是这两局的所有可能结果为

$$(+,+),(+,-),(-,+),(-,-)\text{，}$$

其中使德・梅雷获胜（即至少有一个“+”的情形）有3种，而使保罗获胜（至少有两个“-”的情形）有一种，故德・梅雷获胜的概率为3/4，保罗胜的概率为1/4。这样，德・梅雷应得全部赌金的3/4，而保罗则应得1/4，这就告诉我们，德・梅雷的分法是对的，计算机上的模拟试验也证实了这一点。

他们对这一问题的看法和计算方法不一致，为此而争论不休。后来德・梅雷把这个问题告诉了帕斯卡，帕斯卡对此也很感兴趣，又写信告诉了费马。于是在这两位伟大的法国数学家之间开始了具有划时代意义的通信(帕斯卡最初写给费马的信已不幸丢失)。在通信中，两人用不同的方法正确地解决了这个问题。

下面是费马写给帕斯卡的信(1654)：

“先生：

如果两人赌博时以掷8次骰子为一局，而在下赌注之后我与对方商定，我放弃掷第一次的机会，那么根据我的理论应该得到全部赌金的$\frac{1}{6}$作为补偿。

如果我继续放弃掷第二次的机会，就应得到所剩赌金的$\frac{1}{6}$，即全部赌金的$\frac{5}{36}$作为补偿。

如果第三次轮到我的时候，我仍然弃权，应该得到上次所剩赌金的$\frac{1}{6}$，即全部赌金的$\frac{25}{216}$作为补偿。

如果我第四次弃权，就应得到第三次所剩赌金的$\frac{1}{6}$，即全部赌金的$\frac{125}{1296}$；您说这就是假定掷了前3次之后第四次掷的价值，我完全同意。

但您在来信的最后一个例子中说，如果我在赌博时(以掷8次为一局)要的是6点，而连掷3次都没有得到这个点数。对手建议我不掷第四次，那我就该得到全部赌金的$\frac{125}{1296}$作为补偿。按照我的理论，并非如此。因为在这种情况下，先前掷的3次什么也没得到，赌金总数未变，持有骰子而放弃第四次的人应得全部赌金的$\frac{1}{6}$作为补偿。

如果他已掷4次而没有发现期望的点数，双方商定他不再掷第五次，他依然应得全部赌金的$\frac{1}{6}$，因为赌金总数依然如故。不管是从理论上还是从常识上来说，掷每一次的价值应该是等价的。

我急于知道，您是否同意我的理论，请来信赐教。我相信我们会取得一致的意见，或者仅仅在它的应用方面有些异议。

顺致衷心祝福。

费马 1654”

在 1654 年 7 月 29 日，帕斯卡写给费马的信中，他提到了这个问题和可能的解决方法，

“你的解法非常正确，是给我印象最深的一个，但这些组合太过麻烦。我发现了另一种更为简洁的实在可行的解法”。他在信中写到：

“1.我与你的急切心情是一样的，虽然还卧病在床，但抑制不住要告诉你，我昨晚上从卡尔卡维手里接到了您关于点数问题的来信，我简直不知道用什么语言称赞这封信。我无暇详述，但可用一句话来概括，就是你已发现了如何在两个掷骰子的赌徒之间分配赌徒之间分配赌金的完善方法。看了你那令人信服的论述，我不再怀疑我犯了一个错误。对这一收获，我感到十分满意。

我认为。你关于点数问题的论述比关于骰子问题的论述更值得称赞。我已见过梅雷和罗贝瓦尔等几位先生关于骰子的问题的解答，是梅雷先生向我提起这一问题的。但是他从未发现点数问题的真正价值，也没有找到推导的方法，所以我认为我是唯一知道这种比例的人。

2.您的方法是正确的，而且是我所知道这类问题研究中的首次正确答案。但由于在组合方面会遇到过多的麻烦，我找到另一种更加简洁的方法。我乐于在这里向你作一简单的介绍，因为我很希望在我们今后的讨论中开诚布公。若能取得一致意见，那将是十分愉快的事。我清楚地看到，无论在图卢斯还是在巴黎，真理都是唯一的。

下面给出在两个赌徒之间分配赌金的方法。例如没人放 32 枚金币作为赌金，并以先得 3 分为赢。

假设第一个人已得 2 分，另一个人只有 1 分。他们掷下一次时，若第一个人赢了，她将得到全部 64 枚金币；若另一个人赢了，他们的比分是 2 ：2。如果在这种情况下分赌金的话，每人将拿回自己所下的赌金既 32 枚金币。

综上所述，第一个人如果赢了，64 枚金币将属于他；如果输了，32 枚金币将属于他。假如他们将不希望继续玩下去而要分赌金的话，我一个人应该说：‘我一定能得到 32 枚金币即使我下一轮输了，也应该把它们给我。至于另外的 32 枚金币，也许你得到他们也许我得到他们，机会是均等的，所以，在给我 32 枚金币后，再让我们均分另外的 32 枚吧。’这样，她将得到 48 枚金币，而另一个人只能得到 16 枚。

现在假定第一个人得两分而另一个人得 0 分，他们正在争夺下一分。如果第一个人赢了，他将得到全部 64 枚金币。如果另一个人赢了，注意他们将回到前面的情况，即第一个人有 2 分而另一个人有 1 分。

但我们已说明了在这种情况下已有 2 分的人将得到 48 枚金币。所以，如果他们不希望继续赌下去的话，这人应该说：“如果我赢了，我将得到全部 64 枚金币；如果我输了，48 枚金币将属于我。所以，请先把 48 枚金币给我，然后再均分这剩下的 16 枚金币，因为你我赢得它的机会是均等的”。

于是，他获得金币的数目为

$$48\text{枚}+8\text{枚}=56\text{枚}.$$

如果假定第一个人有 1 分而另一个人为 0 分。先生请看，如果他们再掷一次而第一个人赢了，他与对手的比分将是 2 比 0。根据前述理由，56 枚金币将属于他。如果他输了，他们的比分就成为 1 比 1，32 枚金币将属于他。所以他应该说：‘如果你不打算赌下去，就请把我原来的 32 枚金币给我，再让我们把 56 枚金币的剩余部分均分。56 减去 32 为 24，让我们来平分这

24枚金币吧。你拿12枚我拿12枚，再加上我原来的32枚，我一共应得44枚金币”。

请看，在这种情况下，通过简单的减法，就可以知道如果他赢了第一轮，他将从对方得到12枚金币；如果他继续赢得第二轮，将再从对方得到12枚金币；赢得第三轮则得到8枚金币。

当然，我们还是别把这个问题搞得太神秘了，因为您总是希望问题明朗化的。实际上，我的目的是讨论下述观点的正确性；玩两轮的末轮价值（指从对手赌金中所得数目）是玩三轮的末轮价值的2倍，是玩四轮的末轮价值的4倍，是玩五轮的末轮价值的8倍，等。

3.但是掷末次之前的各次价值的比例就不容易发现了。我考虑了各种情况，终于找到了解决这类问题的方法。我想毫不掩饰地告诉您：当某人赢了第一轮后，不管他希望再掷多少次，都可发现掷第一次的价值。

例如，双方约定再掷8次。前8个偶数和前8个奇数为

2，4，6，8，10，12，14，16和1，3，5，7，9，11，13，15，

用下述方法乘偶数：第一个乘第二个，它们的积乘第三个，再乘第四个、第五个，依此类推。用同样的方法乘奇数：第一个乘第二个，它们的积乘第三个，等。

4.对于任意几个字母，如8个：

$$A, B, C, D, E, F, G, H$$

您可以求出从8个字母中每次取4个的组合数，以及每次取5个字母、6个字母、7个字母、8个字母的组合数。这样将得到所有可能的组合。如果您把每次取4个字母的组合数的一半与更高的组合数相加，其和将等于以2为首项，以4为公比的等比级数的第四项，项数4恰为8的一半。

作为例子，我可用拉丁字母或别的字母表示，因为法文并没什么特别的好处。

例如对任意给定的8个字母

$$A, B, C, D, E, F, G, H。$$

求出它们的所有组合数——每次取4个，取5个，直到取8个。每次取4个的组合数的一半为35（70的一半），把它与每次取5个字母的组合数（56）相加，再与每次取7个字母的组合数（8）相加，最后与每次取8个字母的组合数（1）相加，其和等于以2为首项、以4为公比的等比级数的第4项。项数4为8的一半。

以2为首项、以4为公比的等比级数的各项为 2，8，32，128，512 等。

在这个级数中，第一项是2，第二项是8，第三项是32，第四项是128。128显然等于

35（每次取4个字母的组合数的一半）
+56（每次取5个字母的组合数）
+28（每次取6个字母的组合数）
+8（每次取7个字母的组合数）
+1（每次取8个字母的组合数）。

5.这就是我提出的第一个定理，它纯粹是一个算术问题。关于点数问题的其他有关理论如下：

假定以5轮为一局而某人已得1分，还差4分，整个赌博的输赢将由8次掷骰子决定，而8恰是4的2倍。

如果他在得1分后，放弃掷以后各次的机会，那么他除了拿走自己所下的赌金外，还可得到对手的一部分，其分子是每次从8中取4的组合数的一半（我取4是由于它与失去

的机会相等，取 8 是由于它是 4 的 2 倍），分母是分子与所有更高的组合数之后。

这样，如果我已得 5 分中的 1 分，对手赌金中的$\frac{35}{128}$就归我了。也就是说，如果他下 128 个金币的赌注，我将得到其中的 35 个，而把余下的 93 个留给他。

但是分数$\frac{35}{128}$等于$\frac{105}{384}$，后者的分母恰是从 2 开始的连续 4 个偶数之积，而分子恰是从 1 开始的连续 4 个奇数之积。

您只要略加思索，就可以毫无疑问地看清这一切，所以我认为不必再作进一步的讨论了。

6.尽管如此，我还是要把我的一个旧表寄给您，我没时间抄它。在讨论问题时，我将参考这一表格。

您将很容易地看到，第一轮的价值等于第二轮，这一方法可用组合方法证明。

您还会发现第一行的数总是增加的，第二行、第三行的数也是这样。

但在此以后，第四行的数便递减了，第五行的数也如此。这是很奇怪的。

如果双方各下 256 个金币的赌注，而我在第 n 轮都得分，对手金币中的如下数目将属于我（表 1–2）。

表 1–2

	6	5	4	3	2	1
第一轮	63	70	80	96	128	256
第二轮	63	70	80	96	128	
第三轮	56	60	64	64		
第四轮	42	40	32			
第五轮	24	16				
第六轮	8					

如果双方各下 256 个金币的赌注，而我在前 n 轮得分，对手金币中的如下数目将属于我（表 1–3）。

表 1–3

	6	5	4	3	2	1
第一轮	63	70	80	96	128	256
第二轮	126	140	160	192	256	
第三轮	182	200	224	256		
第四轮	224	240	256			
第五轮	248	256				
第六轮	256					

7.我没有时间在这里写出来麻烦的证明，这个证明曾使梅雷先生大吃一惊，因为他虽有才华但毕竟不是一个几何学家（您知道这是一个很大的缺陷）。他甚至连一条数学上的直线可以无限分割也不理解，他认为直线是由有限个点构成的。我一直没有改变他的看法。如果您能做到这一点，将使他变得完美一些。

………

在我即将完成的一篇几何论文中，我将把上述理论条理化，我已为这篇论文下了一些功夫。

8.在这个题目下，我也作了些算术方面的工作，请不吝指教。

我首先提出一个人人可以接受的引理：从1开始的任意多个自然数所组成的连续级数（如1，2，3，4）中任何两个相邻的乘积，等于较小数与它之前各数之和的2倍。例如

$$4\times5=2\times(1+2+3+4),$$

这就是说没，A 与小于它的各自然数之和的2倍，等于

$$A\times(A+1).$$

现在给出定理：

如果从两个相邻自然数的立方差中减去1，如果等于较小数中所含各数（即从1至该数的所有自然数）之和的6倍。

令 R 和 S 是相差1的两个自然数，且 S 小于 R，则 R^3-S^3-1 等于 S 中所含各数之和的6倍。

令 S 为 A，则 $R=A+1$，R 或 $A+1$ 的立方为

$$A^3+3A^2+3A+1$$

S 或 A 的立方为 A^3，R 与 S 的立方差为 R^3-S^3，所以

$$A^3+3A^2+3A+1-A^3=R^3-S^3,$$

两边减去单位1，得

$$3A^2+3A=R^3-S^3-1.$$

根据引理，把 A 或 S 中所含各数之和加倍，等于 $A\times(A+1)$，即

$$A^2+A,$$

所以 A 中所含各数之和的6倍等于 $3A^2+3A$，但

$$3A^2+3A=R^3-S^3-1,$$

所以 R^3-S^3-1 等于 A 或 S 中所含各数之和的6倍。

没有人怀疑上述证明，但人们说他们并不这样做，因为人人习惯于现成的方法。对我本人来说，把它证明出来也无利可图，只是认为应该承认这是一个很好的证明。尽管如此，我还是期待着您的不同意见。我写出证明的目的就在于此。

9.这里还有两个更困难的问题。我通过一条直线的立方与另一直线的立方相比较，证明了一个平面定理。我认为这是一个纯几何问题，而且是非常精确的，我通过这种方法解

决了下述问题：‘任意给定四个平面、四个点或四个球，作一个球使之给定的球相切，经过给定的点，或经过四个平面围成的四面体的顶点。’还解决了下述问题：‘任意给定三个圆、三个点或三条直线，作一个圆使之与各圆相切，经过各点或与三条直线围成的直线外接。’

我在一个平面上仅仅用圆和直线解决了这些问题，但在证明中使用了立体轨迹——抛物线和双曲线。尽管如此，由于图形建立在平面上，我坚持认为我的解是平面解。

真对不起，我的这封常信打扰了您，使你为我的叙述花了许多时间。我想，我们之间是无话不谈的，所以把自己的见解和盘托出。但愿您能知道我的内心对你是多么的崇敬。祝我们之间的友谊不断增长。”

在 1654 年 10 月 21 日他写给费马的信中提到，当他们互不赞同的时候，能这样通信，保持一致是鼓舞人心的。

他说：“先生，您的最后一封信让我非常满意，您有关‘点数问题’的解法我很钦佩。更是因为我非常理解它完全是属于你的，它与我的解法完全不同，然而却轻易地得到了同样的结果，现在我们又开始和睦了”。

在 1654 年 7 月和 10 月的通信中，他们还联系“点数问题”思考了其他的问题，比如当两人的技艺不等时，或超过 2 人参加游戏的赌金的分配问题。尤其是帕斯卡的研究更有效地推动了数学概率理论的发展，他的组合方法具有一般性。他的工作中还蕴涵了概率论中另一重要的思想——数学期望的思想。在十七世纪弥漫着浓重的宗教神学气质的精神环境中，身为神学家的帕斯卡也结合了宗教和数学两种思想于概率的思考中。

帕斯卡在他的著作《思想录》中曾经提出一个以“帕斯卡赌注”闻名的问题：

“我们既不知道上帝的存在，也不知道上帝的本质。然而我们将倾向于哪一边呢？……，这里进行的是一场赌博，…… 让我们来权衡一下在上帝存在的赌注中的得失。让我们估计这两种可能性，如果你赢了，你赢得所有；如果你输了，你却一无所失。因此，你就不必迟疑去赌上帝的存在吧。”

这个论述中已包含了比较明确的数学期望的思想，这种思想成为以后惠更斯（C.Huggens）和维特（D.Witt）的概率论工作中的一个基本思想，并在以后相当长的时间里在古典概率论的研究中起着重要的作用。

在解决赌本分配问题上，帕斯卡首先接受了卡尔达诺的观点，即公平地分配赌本的原则只与双方为获胜所需赢得的局数 a,b 有关。设 $e(a,b)$ 为 A 最终取胜的概率，总赌本为 1，则 $e(a,b)$ 既是 A 获胜的概率，又是期望，同时还是 A 应获得的总赌本的比例。为了得到 $e(a,b)$，帕斯卡考察了几个较简单的情况（表 1–4）。

表 1–4

	a	b	$e(a-1,b)$	$e(a,b-1)$	$e(a,b)$
(1)	0	$n>0$			1
(2)	n	n			1/2
(3)	1	2	1	1/2	3/4
(4)	1	3	1	3/4	7/8
(5)	1	4	1	7/8	15/16
(6)	2	3	7/8	1/2	11/16

通过细心观察，帕斯卡很快发现其中的规律，并认为对一般情况也成立，即

$$e(a,b)=\frac{1}{2}e(a-1,b)+\frac{1}{2}e(a,b-1)\,, \tag{1-1}$$

边界条件为

$$e(0,n)=1, e(n,n)=1/2, e(a,0)=0$$

上式是一个最简单的偏微分方程，也是一个递归方程。它非常类似算术三角形的一个性质

$$C_{n+1}^{k}=C_{n}^{k}+C_{n}^{k-1}$$

实际上，令

$$e(a,b)=C_{a+b}^{a}=C_{a+b}^{b}\,,$$

则由

$$C_{a+b}^{a}=C_{a+b-1}^{a}+C_{a+b-1}^{a-1}=C_{a+b-1}^{b-1}+C_{a+b-1}^{a-1}\,,$$

可知

$$e(a,b)=e(a-1,b)+e(a,b-1)\,,$$

这与式 1-1 仅差一个 1/2 的因子。帕斯卡对算术三角形的性质当然非常熟悉，因此他猜出

$$e(a,b)=\sum_{i=0}^{b-1}C_{a+b-1}^{i}(\frac{1}{2})^{a+b-1}\,, \tag{1-2}$$

并用证明他的猜想。实际上， 式 1-1 是一个全概率公式的简单应用，即如果赌博不被打断，双方再赌一局，则 A 以 1/2 的概率或胜或败，因此式 1-1 成立。

同样由于信件的遗失，我们也无法得知费马对帕斯卡解法的评价以及他自己的解法。但我们可以从帕斯卡 1654 年 8 月 24 日给费马的回信和 1654 年 9 月 25 日费马给帕斯卡的信中窥测到费马的解法。

费马考虑了一个简单的情况，$(a,b)=(2,3)$。虽然赌博最多需进行 2+3−1=4 局才结束，但也可能在这之前就结束。不过费马认为想象赌博共进行 4 局并不影响赌金的分配结果，这一点可从表 1-5 中看出。

表 1-5

第一局获胜者	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B
第二局获胜者	A	A	A	A	B	B	B	B	A	A	A	A	B	B	B	B
第三局获胜者	A	A	B	B	A	A	B	B	A	A	B	B	A	A	B	B
第四局获胜者	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
最终获胜者	A	A	A	A	A	A	A	A	A	A	A	B	A	B	B	B

从表中可以看出，在总共 16 种结果中，有 4 种 A 在第二局胜出；有 4 种 A 在第三局中胜出；有 3 种在第四局胜出。据此费马得出结论

$$e(2,3)=\frac{4+4+3}{16}=\frac{11}{16}.$$

推广到一般情况，不妨设 $a<b$，则 A 可能在第 a 局取得最终胜利，也可能在第 $a+1,\cdots,a+i,\cdots a+b-1$ 局取得最终胜利。利用加法公式和负二项分布不难得出结论

$$e(a,b)=\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}\left(\frac{1}{2}\right)^{a+i}. \qquad (1\text{-}3)$$

下面证明式 1-2 与 1-3 在形式上不同，但它们实际上是相同的。

为了证明

$$\sum_{i=0}^{b-1}C_{a+b-1}^{i}\left(\frac{1}{2}\right)^{a+b-1}=\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}\left(\frac{1}{2}\right)^{a+i},$$

我们首先给出一个引理

$$\sum_{k=0}^{n}C_{r+k}^{k}C_{m-k}^{n-k}=C_{m+r+1}^{n}$$

这个引理的证明的思路是这样的：考虑一个人 A 从原点(0, 0)到点 $(m-n+r-1,n)$ 最短路径的个数。

下面证明式 1-2 与 1-3 相等。

显然

$$\sum_{i=0}^{b-1}C_{a+b-1}^{i}\left(\frac{1}{2}\right)^{a+b-1}=\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}\left(\frac{1}{2}\right)^{a+i},$$

等价于

$$\sum_{i=0}^{b-1}C_{a+b-1}^{i}\left(\frac{1}{2}\right)^{a+b-1}=\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}2^{(b-1)-i}$$

利用二项式展开定理及交换求和次序我们有

$$\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}2^{(b-1)-i}=\sum_{i=0}^{b-1}C_{a-1+i}^{a-1}\sum_{j=0}^{(b-1)-i}C_{(b-1)-i}^{j}=\sum_{j=0}^{b-1}\left[\sum_{i=0}^{(b-1)-j}C_{a-1+i}^{a-1}C_{(b-1-i)}^{j}\right],$$

由引理知

$$\sum_{i=0}^{(b-1)-j}C_{a-1+i}^{a-1}C_{(b-1-i)}^{j}=C_{a+b-1}^{(b-1)-j},$$

所以

$$\sum_{j=0}^{b-1}\left[\sum_{i=0}^{(b-1)-j}C_{a-1+i}^{a-1}C_{(b-1-i)}^{j}\right]=\sum_{j=0}^{b-1}C_{a+b-1}^{(b-1)-j}=\sum_{i=0}^{b-1}C_{a+b-1}^{i},$$

因此， 式 1-2 与 1-3 相等。

费马并未直接给出式 1-3，但他在 9 月 25 日的信中明显认为 $(a,b)=(2,3)$ 时的结论对一般情况也成立。

至此，帕斯卡与费马终于联手解决了赌本分配问题，同时也为最终建立概率原理迈出了伟大的一步。不过费马明显使用的是负二项式，而现代人却称之为帕斯卡分布。

帕斯卡和费马正确解决了“点数问题”的这一事件被伊夫斯（Howard Eves）称为“数

学史上的一个里程碑”。在概率论的历史上，一般的传统观点则把这一事件看作为数学概率论的起始标志。之所以不把卡尔达诺的著作作为概率论的起源的始点，有这样几个原因：在卡尔达诺的著作中只有一小部分内容是处理机会（chance）的计算的。就像卡尔达诺的大多数作品一样，这种处理似乎只是零碎的和模糊的，混杂于卡尔达诺的个人的一些奇闻轶事、哲学思考、大量流行的赌博者常用的欺骗策略和精明的心理应用等建议之中，并且他的这本著作中所阐述的数学思想对数学家和一般的赌徒几乎都没有什么影响。因为对于当时的数学家而言，概率太游戏化了，而对赌徒而言，概率又太数学化了。而帕斯卡和费马的通信除了正确解决了一些问题和概念之外，还创造了一种研究的传统——用数学方法（主要是组合数学的方法）研究和思考机会性游戏。这种传统统治这个领域达半个多世纪的时间。所以，综合考虑所有这些因素，这个事件赢得它在数学概率论的历史中的标志性的地位是当之无愧的。

1.4 惠更斯的《论赌博中的计算》研究

惠更斯（Christian Huygens 1629~1695）于1629年4月14日诞生于海牙的一个富豪之家。他的父亲是一个杰出的诗人和外交家。惠更斯从小就喜欢钻研学问，跟随父亲学习了数学和力学。十六时，惠更斯进入莱顿大学，后转到布雷达大学学习法律和数学。1650年起，惠更斯开始研究光学，同时对天文观测产生了浓厚兴趣。1655年获得法学博士学位后，惠更斯转入科学研究。他先后访问了伦敦和巴黎，并在巴黎获得了普遍的尊敬。1663年，惠更斯成为英国皇家学会第一个外国会员和并被巴黎科学院接纳为唯一的一个外国院士。在伦敦和巴黎时，惠更斯结识了许多当时著名的科学家，包括牛顿、莱布尼兹等。在巴黎生活的第十五年，法国和荷兰之间爆发了战争，惠更斯不得不离开巴黎，回到故乡荷兰，过着孤独寂寞的晚年生活。1695年6月8日，惠更斯在海牙逝世。

克里斯蒂安·惠更斯是与牛顿同一时代的科学家，是历史上最著名的物理学家之一，他对力学的发展（特别是摆的研究和运用）和光学的研究都有杰出的贡献，在数学和天文学方面也有卓越的成就，是近代自然科学的一位重要开拓者。

惠更斯在数学上有出众的天才，早在22岁时就发表过关于计算圆周长、椭圆弧及双曲线的著作。他对各种平面曲线，如悬链线、曳物线、对数螺线等都进行过研究，还在概率论和微积分方面有所成就。

在纪元之初，民间就流行用抽签来解决人们彼此间的争端，这可能是最早的概率应用。随着社会的发展，随机现象愈来愈左右着人类的生活。因而在不确定性因素的情境中，寻找行为的理性规则，使理性服从机遇的愿望成为数学家研究的课题之一。直到文艺复兴时期，随机世界依然扑朔迷离、不能辨析。作为研究随机现象的概率论出现在17世纪中叶，象征着概率论诞生的标志，就是克里斯蒂安·惠更斯在1657 年发表的《论赌博中的计算》(On Reckoning at Games of Chance) 一文。

1.论文的来源

惠更斯1629年诞生于海牙的一个富豪之家。其父知识渊博，擅长数学研究，同时又是一杰出的诗人和外交家。惠更斯从小受到了父亲的熏陶，喜欢学习和钻研科学问题。16岁进入莱顿大学学习，后转到布雷达大学学习法律和数学。26岁获得法学博士学位。数学

老师范·舒藤(Frans Van Sehooten) 指导他学习当时的著名数学家、哲学家卡卡维(Carcavi)的数学著作及其哲学著作。惠更斯从中感悟到数学的奥妙而对数学很感兴趣。1650~1666年期间，他大多时间在家中潜心研究光学、天文学、物理学和数学等领域，成果显著，一举成为当时闻名遐迩的科学家。除去在光学、天文学等领域的贡献外，惠更斯也有出众的数学才能，可谓是一个解题大师。早在 22 岁时就写出关于计算圆周长、椭圆弧及双曲线的论文。他发现了许多数学技巧，解决了大量数学问题。如他改进了计算 π 值的经典方法；继续笛卡尔、费马和帕斯卡的工作，对多种平面曲线，如悬链线、曳物线、对数螺线、旋轮线等都进行过研究；对许多特殊函数求得其面积、体积、重心及曲率半径等，某些方法与积分方程的积分法相似。伯努利兄弟对惠更斯的研究极为佩服，尤其是约翰(John Bernoulli，1667~1748)发现旋轮线也是最速降线时甚是激动。他说："这惠更斯等时曲线(旋轮线) 就是我们正在寻求的最速降线！我感到十分惊奇！"

惠更斯在数学方面的最大贡献，就是以《论赌博中的计算》一文奠基了概率论的基础。1654 年，赌徒德·梅雷向当时的"数学神童"帕斯卡提出了其在赌场上遇到的几个不解问题。后帕斯卡与费马以通信的方式对这些问题进行了较为详尽的讨论，并将其推广到一般情形，这就使概率计算由单纯计数而转向更为精确的阶段，但二人都不愿意发表研究成果，故有关概率知识没有得到及时传播。1655 年秋，惠更斯第一次访问巴黎。他遇到罗贝瓦尔(G.P.de Roberval)及梅雷恩(Mylon)，但没有见到帕斯卡和费马。他获知去年有一场关于概率问题的讨论，但不知其具体解决方法及结果。由于罗贝瓦尔对此问题毫无兴趣，因而惠更斯对费马和帕斯卡的讨论结果几乎一无所知。1657 年 3 月在最后一次校订时，惠更斯将其论文增加为 9 个命题和 5 个问题，形成了《论赌博中的计算》1656 年 4 月，回国后的惠更斯自己解决了这些概率问题，并将其手稿送给范·舒藤审阅，同时写信给罗贝瓦尔，寻求几个概率问题的解答。此时范·舒藤正在筹印其《数学习题集》，因而他建议惠更斯将此文印刷发表，并亲自替学生将该文译成拉丁文。由于惠更斯没有收到罗贝瓦尔的信，便又写信给梅雷恩，并通过卡卡维将信转给费马。在 1656 年 6 月 22 日费马的回信中，给出与惠更斯相一致的解决方案，但无证明过程。此外，费马又向惠更斯提出了 5 个概率问题。阅信后，惠更斯很快解出这些问题，并把其中 2 个问题收录在著作中。他于 7 月 6 日将结果送给卡卡维让其转给梅雷恩、并给出帕斯卡与费马对点数问题的解决方案帕斯卡和费马确定解答正确与否。卡卡维在日给惠更斯的回信中也提出了一个无证明的解决方法。但无证明。惠更斯在 9 月 28 日的回信中肯定了卡卡维的解答，10 月 12 的基本构架。惠更斯还将给范·舒藤的一封信作为该文的前言，这篇前言形成了全文的思想基础。他在其中明确地提出：尽管在一个纯粹运气的游戏中结果是不确定的，但一个游戏者或赢或输的可能性却可以确定。"可能性"用的是"probability"，其意义与今天的概率几无差别。惠更斯的这种思想使得"可能性"成为可以度量、可以计算' "我相信，只要仔细具有客观实际意义的概念。信中惠更斯强调了这一新理论的重要性：研究这个课题，就会发现它不仅与游戏有关，而且蕴涵着有趣而深刻的推理原则"。"并惋惜地说，法国的杰出数学家已经解决了这些问题，无人会把这个发明权授予给我"。其内容被编排在范 舒藤之书的 519~534 页。该书出版于 1657 年 9 月，而荷兰文版出版于 1660 年，英文版出版于 1692 年，德文版出版于 1899 年，法文版出版于 1920 年，意大利文版出版于 1984 年。

1657 年发表的《论赌博中的计算》，就是一篇关于概率论的科学论文，显示了他在数

学上的造诣。卡尔达诺的工作对 17 世纪的概率研究没有什么影响，因为当人们知晓他在概率方面的成就时已经太迟了。在 1657 年，惠更斯的关于概率论的小册子业已公之于世。

惠更斯对古典概率进行研究，以及他后来取得超出同辈的成就，其起因就在于他听到帕斯卡和费马在进行这方面的研究。如果没有 1655 年的巴黎之行，在概率论这一数学领域，他可能永远不会有所建树。1655 年 7 月至 12 月惠更斯在巴黎期间，结识了不少法国学术界名流，并由此了解到了当时法国的学术动态。尽管他未能幸会费马及帕斯卡(这是因为费马那时远在图鲁斯((Toulous)，帕斯卡虽在巴黎，但由于信仰及宗教上的迷茫而拒见任何人)，但是，惠更斯有幸从米隆(Mylon)及罗伯弗尔那儿得知费马与帕斯卡在讨论概率问题，不过仅仅是知道这件事而已。“他没能获得他想知道的费马与帕斯卡对概率所做研究的详细情况”。所以，能够肯定，虽然惠更斯涉及的概率问题，在内容上没有超出帕斯卡及费马的研究范围，但他确实是独立地建立起自己的理论的，而且还可以看到，惠更斯提出的方法，要比帕斯卡及费马的方法优越。帕斯卡和费马仅解决了一些零散的概率问题，惠更斯则从概率的——期望值出发，为经典概率建立了系统的理论。

概率论学科的形成归功于惠更斯，1655 年他访问巴黎期间了解到帕斯卡和费马的通信，对概率问题产生了兴趣，次年写成一篇论文《论赌博中的计算》，并于 1657 年作为其老师范·舒腾的《数学练习》的最后一卷发表。虽然惠更斯讨论的仍是骰子赌博问题及其计算，但是与前人不同，这篇论文数学特色突出，具有科学著作的规范性，赌博是理论的模型，或者是一类问题，而不是论文的全部意义。

论文的结构是一个引言、一个公设、十四个命题和一个推论以及五个供读者练习的问题。引言指出“虽然在一个纯粹运气的游戏中结果是不确定的，但一个游戏者或赢或输的可能性可以确切地确定。”可能性用的“probability”，意义与今天的概率没有差别。惠更斯的这种认识使得“可能性”真正成为可以度量、具有客观实在意义的概念了。

2.创立数学期望

《论赌博中的计算》的写作方式很像一篇现代的概率论论文。先从关于公平赌博值的一条公理出发，推导出有关数学期望的三个基本定理，利用这些定理和递推公式，解决了点数问题及其他一些博弈问题。最后提出 5 个问题留给读者解答，并仅给出其中的 3 个答案。通常所谓惠更斯的 14 个命题，指的就是书中 3 条定理加上 11 个问题。

公理：每个公平博弈的参与者愿意拿出经过计算的公平赌注冒险而不愿拿出更多的数量。即赌徒愿意押的赌注不大于其获得赌金的数学期望数。

对这一公理至今仍有争议。所谓公平赌注的数额并不清楚，它受许多因素的影响。但惠更斯由此所得关于数学期望的 3 个命题具有重要意义。这是数学期望第一次被提出，由于当时概率的概念还不明确，后被拉普拉斯(P.S.Laplace，1749~1827) 用数学期望来定义古典概率。在概率论的现代表述中，概率是基本概念，数学期望则是二级概念，但在历史发展过程中却顺序相反。

公设提出数学期望的概念：赢取某物的机会或期望(Chance or Expectation)等于这样一个和，即是在一个公平赌博中他将以同样的机会和期望会获得的那些。虽然措辞有点晦涩，不过他以赌博情形作了解释：如果一个人将 3 先令放于一只手而将 7 先令放于令一只手，让我选择其中之一，我说这与他给我 5 先令是一样的。实际上惠更斯的“期望”就是帕斯卡的“机会的值”(value of chance)。

从公设出发，惠更斯首先巧妙地证明了三个命题，它们是其余命题的基础。

关于数学期望的三个命题为：

命题 1：如果赢取 a 及 b 的机会相等，那么整个赢面为

$$(a+b)/2$$

命题 2：如果赢得 a，b 及 c 的机会均等，那么整个赢面为

$$(a+b+c)/3$$

命题 3：如果赢取 a 的机会是 p，赢取 b 的机会是 q，那么整个赢面为

$$(pa+qb)/(p+q)$$

这些今天看来都可作为数学期望定义。但对惠更斯来说，必须给出演绎证明，因当时对数学的一种公认处理方法是从尽可能少的公理推导其他内容。惠更斯所给的命题 1 证明为：

假设在一公平的赌博中，胜者愿意拿出部分赌金分给输者。若二人的赌注均为 x，胜者给输者的为 a，因而所剩赌金为 $2x-a=b$，故

$$x=(a+b)/2$$

帕斯卡与费马在通信中所说的“值”等于赌注乘以获胜的概率，因而已与概率无本质区别。而惠更斯在这里将“值”改称为“数学期望”是一个进步(在该书荷兰版中，惠更斯仍沿用“值”的概念)。

将命题 3 推广便得到今日数学期望的定义。因此惠更斯当之无愧是数学期望概念的奠基人。

可以认为，这 3 个命题蕴涵了惠更斯的概率理论。第一个命题是说，两人赌博，把所下的赌注分成 a，b 两份，并规定胜者得 a，负者得 b。在公平赌博中，两人的胜负机会是相等的，因此，他们得到 a 或 b 的概率都为 1/2。第二个命题不过是把参加赌博的人数由两个增至 3 个。重要的则是命题 3，因为它把经典概率问题完全包容了。对命题 3，可以如下解释。假定有 $p+q$ 个参加赌博的人，他们有均等的胜负机会。令 $p+q=r$，用 $A_1,A_2,\cdots,A_r$，代表这些赌博者，并使他们环桌而坐，且令每个人下的赌注为 $(pa+qb)/r$。规定如果 A_i 获胜，那么他本人得 b，且在他左边的 $q-1$ 人中的每个人，也得 b；而 A_i 右边的 p 个人，每人只能得 a。从而，每个参加赌博的人得到 a 的概率为 p/r，得到 b 的概率为 q/r。因此，可以推出，每个人的赢面是

$$(pa+qb)/(p+q)$$

在该命题中，惠更斯论证的是，对任一参加赌博的人，不论赌博方式如何，也不论有多少人参加，只要他获取 A_i 的概率为 p_i，那么他的总赢面就是

$$N=\sum_{i=1}^{r} P_i A_i$$

此公式概括了惠更斯在概率方面的全部工作，它给出了古典概率问题的一般求解方法，这是帕斯卡与费马所未做到的。

3.求解点数问题

所谓点数问题是：甲乙二人赌博，其技巧相当，约定谁先胜 s 局则获全部赌金。若进行到甲胜 s_1 局而乙胜 s_2 局时($s_1<s$，$s_2<s$)，因故停止，赌金应如何分配才公平？惠更

斯深刻认识到点数问题的重要性，因而在其著作中有 6 个命题讨论了该问题。命题 4~7 都是有关二人的点数问题，而命题 8 和命题 9 将问题推广到三人及若干个人。

惠更斯的解决思路为：赌徒分得赌注的比例等于其获胜的概率。他假设赌徒在每局获胜的概率不变，且各局间相互独立。这样就可以归结为一般问题：

设随机试验中某随机事件每次成功的概率为 p，重复独立进行该试验若干次，求在 b 次失败前取得 a 次成功的概率。

惠更斯给出的另外 11 个命题(事实上是问题)，均可用上面公式来求解。命题 4 至命题 8 属于赌金分配问题。

命题 4：假定我与某人赌博，并规定掷出 3 个赢点才为胜。若我已掷出两个赢点，而对家只掷出一个赢点，至此我们决定不再赌下去，那么应如何分赌金?

命题 5：已知甲需再赢 1 点，乙需再赢 3 点，才能胜出，此时如果中止赌博，问甲、乙如何分配赌金?

命题 6：我掷出一个赢点即算胜，对家须掷出 3 个赢点才能获胜，在此情况下应如何分赌金?"

命题 7：已知甲需再赢 2 点，乙需再赢 4 点 ，问甲、乙如何分配赌金?

命题 6，7 无须赘述，它们与命题 4 的不同之处，不过是为获胜而必须掷出的点个数不同而已。

命题 8：甲、乙、丙三人，甲和乙各需再赢 1 点，丙需再赢 2 点。中止赌博时，三人如何分配赌金?

命题 9：若干人赌博，已知各需再赢点数，要计算赌金的分配比例，首先必须考虑各人赢得下一局的情形。

上面两个命题是三人或三人以上的点数问题。剩下的 6 个命题属于用骰子掷出某个点子的概率问题。

命题 10：一个骰子掷几次，掷出 6 点的概率为多少?

命题 11：两个骰子掷几次，掷出双 6 点的概率为多少?

以上两命题相当于求掷得 6 点和 12 点的概率。

命题 12：一次至少能掷得两个 6 点，求骰子数。

4. 独创分析法

在《论赌博中的计算》的最后两个命题中，惠更斯创立了著名的“惠更斯分析法”来解决概率问题。

命题 13：甲、乙掷一对骰子，约定：若掷得 7 点，则甲赢；若掷得 10 点，则乙赢；若掷得其他点数，则平分赌金。问甲、乙二人各自的期望值。

命题 14：用两个骸子赌博，假如我掷出的点子和为 7 点，就可拿走全部赌金；对家掷出的点子和为 6 点时亦可拿走全部赌金。让他先掷，那么我与他的胜率之比为多少?

惠更斯在讨论这些命题时，应用了期望值这一数学概念。利用上面公式，能够很容易地求出这些问题的结果。就命题 6 而言，对家须掷出 3 个赢点，那么自家的胜率为多大呢?显然，第一个赢点的掷出，双方都有对半的机会，即 $P_1=1/2$。此外，还有 1/2 机会可予期望，因此掷出第二个赢点的概率期望值为 $P_2=1/4$。同理 $P_3=1/8$。从而自家应分的赌金为

$$N=\sum_{i=1}^{r}P_iA=A(P_1+P_2+P_3)=\frac{7}{8}A.$$

这就是说应得全部赌金的 7/8。

惠更斯对命题 10 是这样论述的：

很明显，在第一次投掷中，掷出 6 点的机会是一次，而非 6 点的机会是 5 次，由命题 3 可知，他赢面的期望值为$a/6$，而对家赢面的期望值为$5a/6$。所以，在第一次投掷中，他与对家赢面期望值之比是 1：5。

至于投掷两次，掷出 6 点的概率与非 6 点的概率之比，可按如下方式求出。如果他在第一掷中掷出 6 点，则可拿到a；若掷出的不是 6 点，则他还可以再掷一次。由前面的论证可知，他在第二次投掷中，只有一次机会掷出 6 点，而有 5 次却是非 6 点。也就是说，他有 1/6 的概率赢得 $5a/6$。从而，由命题 3 可得，他的总赢面为 $11a/36$。亦即他得到a的期望概率为 1/6，得到 $5a/6$ 的期望概率为 1/6，因此，总赢面应为$a\cdot\frac{1}{6}+\frac{5a}{6}\cdot\frac{1}{6}=\frac{11}{36}a$。惠更斯用相同的方法，进一步求出掷 3 次骰子掷出 6 点的概率为 91/216；掷 4 次掷出 6 点的概率为 671/1296；掷 5 次掷出 6 点的概率为 4651/7776。令为n次投掷掷出 6 点的概率，则

$$P_n=\frac{1}{6}+\frac{1}{6}\cdot\frac{5}{6}+\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}+\cdots+\frac{1}{6}\cdot(\frac{5}{6})^{n-1}=1-(\frac{5}{6})^n$$

对命题 11，惠更斯所采用的论述方法同前。令P_n为n次投掷掷出双 6 点的概率，且令$V_n=aP_n$。当$n=1$时，掷出双 6 点的概率为 1/36；当$n=2$时，他有 1/36 的概率获得$35a/36$，所以

$$V_2=\frac{a}{36}+\frac{35a}{36}\cdot\frac{1}{36}=\frac{71a}{1296}$$

显然$P_2=\frac{71}{1256}$。当n为更大的数时，惠更斯也进行了论证。

对于$n=4$，在前两次投掷中，掷出双 6 点的概率为P_2，同时$(1-P_2)V_2$可予期望。由命题 3 可得

$$V_4=P_2a+(1-P_2)V_2=178991a/1679616,$$

当$n=8$，16，24 时，惠更斯给出如下公式

$$V_8=P_4a+(1-P_4)V_4,$$
$$V_{16}=P_8a+(1-P_8)V_8,$$
$$V_{24}=P_{12}a+(1-P_{12})V_{12},$$

尽管惠更斯没有给出它们的数值，但他发现P_{24}略小于 1/2，P_{25}略大于 1/2。

命题 14 与命题 11 无多大差别，所不同的是命题 14 涉及非公平赌博，先掷的人较有利。因为，如果他在第一掷中就掷出赢点，那么对家就失去了再掷的机会。对这种问题，惠更斯仍是借助期望值这一工具来解决的。其方法如下："设我的赢面为x，全部赌金为a，那么对家的赢面就是$a-x$。显然，每次轮他掷时，我的赢面为x，但在轮我掷时，我的赢

面会更大一些，假定为 y。现在用两个散子掷 36 次，我的对家有 5 次机会掷出两个点子和为 6 点的结果(即 3+3，4+2，2+4，5+1，1+5)。因此，有 31 次投掷对他是不利的。即使在这 31 次中，轮我掷，我也只有在这 31 次投掷中期望赢到 y，其余 5 次是对家的权利，对我而言是不存在的。由命题 3 可知，赢得 y 的概率为 31/36。由于先前已假定 $31y/36=x$，所以 $y=36x/31$。此外，还假定了轮我投掷时，赢面为 y。我掷骰子时，有 6 次机会得到 a，因为我有 6 次可能掷出两个骰子的点数和为 7 点的结果；另一方两，又有 30 次机会轮对家投掷，此对，我的赢面又变为 x。因而，由命题 3 可知，y 值等于赢得 a 的 6 次机会与赢得荆的 30 次机会之和，即 $(6a+30x)/36$ 是我的赢面，它应等于 y。根据前面的推导可得，y 等于 $36x/31$，因此必有

$$\frac{30x+6a}{36}=\frac{36}{31}x\,,\quad x=\frac{31}{61}a\,.$$

此即为我之赢面，从而对家的赢面就是 $30a/61$。所以，我与对家赢面之比为 31：30。在命题 14 的计算中引入了两个方程，这种方法后来被伯努利称为“惠更斯的分析方法”。

从惠更斯解决概率问题的过程中，可以看到，他深刻地把握住了概率的数学意义，紧紧扣住了期望值这一关键。他的概率理论，或者说，他提出的公式，从原则上讲，是能够解决所有经典概率问题的。

费马与帕斯卡以信函方式来讨论概率问题。从克拉莫(E.E. Kraxner)的《近代数学的本质及发展》(The Nature and Growth of ModernMathematics)一书中可知，他们是用算术三角形来处理概率问题的。就赌金分配问题而言，该方法只不过是穷竭所有的可能性，因而没有一般的期望值这一概念。帕斯卡曾论述过如下一个关于赌金分配的问题：“假定 A，B 两人赌博，A 掷出两个赢点即可得全部赌金，而 B 须掷出 3 个赢点才能得到全部赌金。因此，在 4 次都有赢点掷出的赌局里，有下列可能的情况：

AAAA	*BAAA*	*BBAA*	*BBBA*	*BBBB*
		ABAA	*BABA*	*BBAB*
		AABA	*ABBA*	*BABB*
		AAAB	*BAAB*	*ABBB*
				ABAB
				AABB

……在这些组合中，含有 2 个、3 个或 4 个 A 的项，对 A 是有利的，含有 3 个或 4 个 B 的项对 B 是有利的。因此，A 的赢面为 11，B 的赢面为 5。这就是说，A 与 B 要按 11：5 来分赌金。帕斯卡在概率方面的工作大体如此，尽管有些数学史著作认为“他解决了一般性的赌金分配问题，并利用算术三角形求出了许多这类问题的结果”，但是算术三角形方法不具有普适性。对于 3 人以上的赌金分配问题，该方法并不总是有效的；而对惠更斯的命题 14 所涉及的非公平赌博问题，这种方法就无能为力了。出现这种情况的原因在于算术三角形方法不涉及期望值，帕斯卡本人也不理解这个概念。也正是由于不理解，他才会在讨论骰子概率问题时出现差错。帕斯卡在给费马的信函中有过这样的论述：“如果在 8 次投掷里，我已 3 次未掷出可以赢局的 6 点，若对家建议我放弃第 4 次投掷，并让我拿走

一部分赌金作为放弃该次投掷的补偿，那么我应取走全部赌金的 125/1296。只要学过一些概率就会知道，此处，帕斯卡给的结果是错误的，正确的答案应是 1/6。后来费马回信纠正了帕斯卡的这个错误。

5.惠更斯的 5 个问题

惠更斯的最后 5 个问题，虽也都是在形形色色的赌博机制中，计算一方取胜的概率，其中只有问题 1、3、5 给出来答案，但我们可以在其他文献中查到惠更斯的解答，但在概率论诞生初期，这无疑是向同时代数学家的挑战。

他说："给我的读者(如果有的话) 留下一些思考题应该是有益的，这将供他们练习或者打发时间"。

问题 1： A、B 两人轮流掷两枚骰子，A 先掷，然后 B 连续掷两次，然后 A 再掷两次，以此类推。若 A 掷出 6 点，则 A 胜；若 B 掷出 7 点，则 B 胜。问 A 的赔率是多少?

惠更斯在 1656 年 6 月 6 日给喀卡维的信中对问题 1 进行了解答。

利用前面提到的分析法，令 a 为总赌本，$p_1=1-q_1$，$p_2=1-q_2$ 分别为 A、B 一次掷出 6 点或 7 点的概率，e_n 为在第 n 次投掷前的期望。则由于比赛每 4 局构成一个循环且投掷之前相互独立，因此只需考虑 e_1,e_2,e_3,e_4。惠更斯的推理可以表述成如下 4 个方程

$$e_1=p_1a+q_1e_2,\ e_2=q_2e_3,\ q_3=q_2e_4,\ e_4=p_1a+q_1e_1,$$

由这 4 个方程可得到 A 的期望

$$e_1=\frac{p_1(1-q_1q_2)}{1-q_1^2q_2^2}a,$$

获胜的概率为

$$\frac{e_1}{a}=\frac{p_1(1-q_1q_2)}{1-q_1^2q_2^2}$$

由 P_1=5/36，P_2=6/36 可得 A 的赔率为 10 355 : 12 276。

问题 2： A、B、C 三人轮流抽取 12 张筹码（放回）。筹码中有 4 张白色的，8 张黑色。先抽到白色筹码的人获胜。问 A、B、C 获胜的比率是多少?

惠更斯在 1665 年的一篇短文中给出了问题 2 的解答。

设 x，y，z 分别为 A、B、C 三人的期望。由于是轮流放回式的抽取筹码，因此在 A 第一次未抽取到白色筹码的条件下，A 的期望就等于开始抽取前 C 的期望 z，B 的期望就等于开始抽取前 A 的期望 x，C 的期望就等于开始抽取前 B 的期望 y。所以有

$$x=pa+qz,\ y=qx,\ z=qy,$$

其中 a 为总赌本，$p=1-q=1/3$ 为一次抽取到白色筹码的概率。由此可得

$$x=\frac{p}{1-q^3}a,\ y=\frac{pq}{1-q^3}a,\ z=\frac{pq^2}{1-q^3}a$$

将 $p=1/3$ 代入，可得 A、B、C 三人获胜的比率是 9 : 6 : 4。

问题 3：有四种花色的扑克牌，每种花色有 10 张，共 40 张。A 与 B 打赌：A 从中任取 4 张，若 4 张为不同花色，则 A 获胜。问 A 获胜的概率是多少?

在1656年6月6日给喀卡维的信中，惠更斯未加证明地给出了问题3的答案，一般相信惠更斯的推理如下：

设总赌注为1，且在已经抽取n种花色的扑克牌的条件下，A获胜的期望为e_n，则A获胜，那么下一次抽到的牌必须是$10(4-n)$张牌中的一张，由此有

$$e_n = \frac{10(4-n)}{40-n} e_{n-1},$$

由边界条件$e_4=1$，得到

$$e_0 = (10/37)(20/38)(30/39)(40/40) = 1000/9139$$

问题4：有12张筹码，其中4张白色的，8张黑色的。现A与B打赌：若从中任意抽取的7张筹码中，恰有3张白色4张黑色的，则A获胜。问A获胜的概率是多少?

在1665年的一篇短文中，惠更斯给出了问题4的解答。

设$e(a,b)$为在已知取得a张白色的筹码，b张黑色的筹码的条件下，A获胜的期望。若再从剩下的12−a−b张筹码中任取其一，则有

$$e(a,b)\frac{4-a}{12-a-b}e(a+1,b)+\frac{8-b}{12-a-b}e(a,b+1),$$

其中$0 \le a \le 4$，$0 \le a+b \le 7$。

利用边界条件$e(3,4)=t$（t为总赌本），当a=0,1时，

$$e(a,7-a)=0,\quad e(a,6-a)=0,\quad e(a,5-a)=0,$$

惠更斯经过19步的反复递推得到

$$e(0,0) = \frac{35}{99}t,$$

即A获胜的概率为35/99。

惠更斯还注意到问题4的逆问题，即只取5张筹码，其中1张白色的，4张黑色的，他只用了9步就得到了A获胜的概率为42/99。

问题5：A、B两人各有12个筹码。现掷三枚骰子，若出现11点，则A给B一个筹码；若出现14点，则B给A一个筹码。先将对方筹码赢光者获胜。问A的胜率是多少?

此问题称为赌徒输光问题(Gambler' s Ruin Problem)。这个问题首先由帕斯卡提给费马并通过喀卡维转给惠更斯。惠更斯在1656年10月12日给喀卡维的信中给出了答案。下面的解答来源于1676年惠更斯写的一篇短文。

与帕斯卡一样，惠更斯设A、B两人初始分数为(0，0)，则领先对方12分者获胜。由于掷3枚骰子共有216种结果，其中有27种和为11点，15种和为14点，由此A获得1分的概率为$p=15/42=5/14$，B获得1分的概率为$q=9/14$。

令$e(a,b)$为在A已经获得a分，B获得b分的条件下，A获胜的概率(设总赌本为1，则概率等于期望)，问题就变为求$e(0,0)$。

惠更斯首先考虑了几种简单的情况，领先2分者获胜。惠更斯将所有可能的结果用简单的图1-1表示出来。

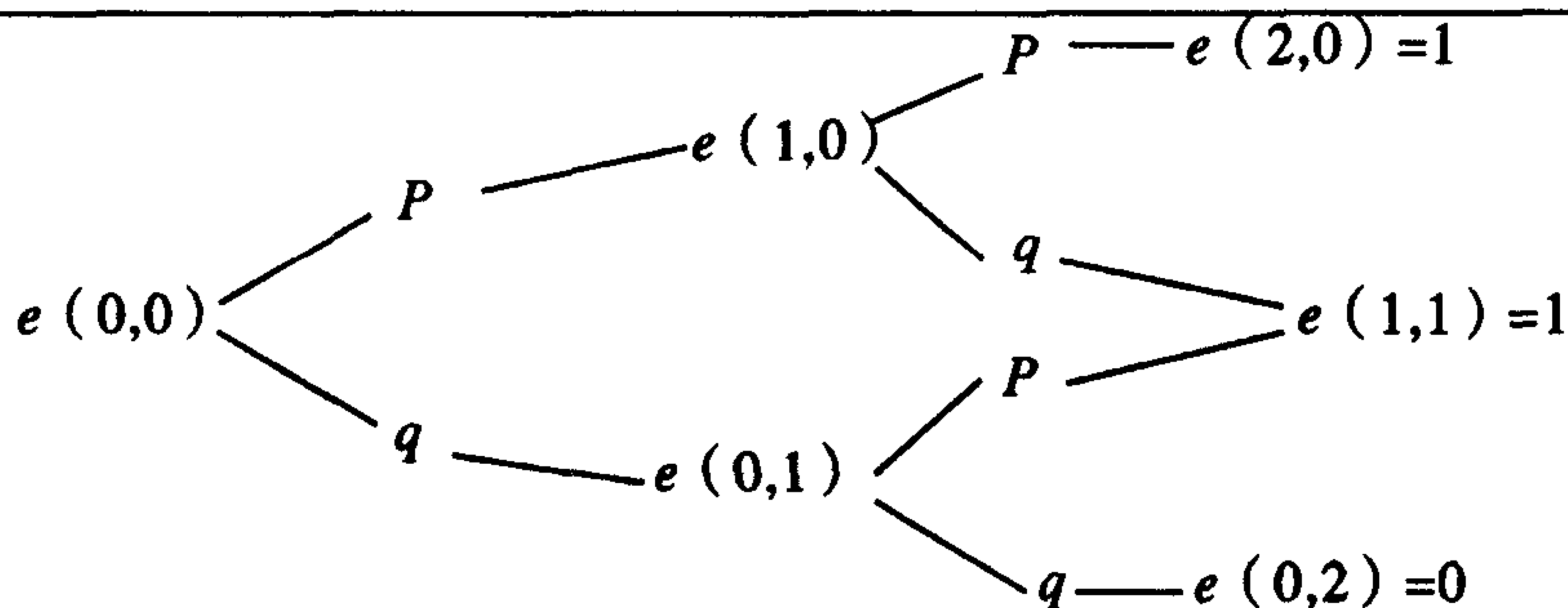


图 1-1

由图 1-1 立刻可以得到如下方程

$$e(1,0)=pe(2,0)+qe(1,1)=p+qe(0,0),$$
$$e(0,1)=pe(1,1)+qe(0,2)=pe(0,0),$$
$$e(0,0)=pe(1,0)+qe(0,1),$$

其中 $e(2,0)=1$，$e(0,2)=0$ 为边界条件。

由上面的方程可解得

$$e(0,0)=\frac{p^2}{p^2+q^2},$$

由此，惠更斯得到领先 4 点或领先 8 点情况下获胜的概率为

$$\frac{p^4}{p^4+q^4}\text{及}\frac{p^8}{p^8+q^8}.$$

为了计算领先 3 点者获胜的概率，惠更斯考虑从(1，0)出发，A 获胜的概率为 $e(1,0)$。如果从(1，0)出发，A 获胜，那么 A 必须再领先 2 分，这个概率已由前面给出，并且由于 A 或 B 领先 2 分为必然事件，因此有

$$e(1,0)=\frac{p^2}{p^2+q^2}e(3,0)+\frac{q^2}{p^2+q^2}e(1,2)=\frac{p^2}{p^2+q^2}+\frac{q^2}{p^2+q^2}e(0,1),$$
$$e(0,1)=\frac{p^2}{p^2+q^2}e(2,1)+\frac{q^2}{p^2+q^2}e(0,3)=\frac{p^2}{p^2+q^2}e(1,0),$$
$$e(0,0)=pe(1,0)+qe(0,1),$$

由此解得

$$e(0,0)=\frac{p^3}{p^3+q^3}.$$

利用领先 2 分获胜和领先 3 分获胜的结论不难得到 A 领先 6 分的概率为

$$\frac{p^6}{p^6+q^6},$$

最终惠更斯表示，在领先 n 分获胜的情况下，　A 与 B 获胜的概率之比为

$$p^n:\ q^n.$$

6.历史评价

到17 世纪时，不少学者已对赌博中的某些问题进行了讨论，并挖掘了其中的数学原理。但对当时的大多数学家来说，概率论是庸俗的赌博游戏，难登大雅之堂。正是社会的发展及其需要，才推动了概率论的发展。如果没有社会的需要，概率论至今恐怕仍然只能在牌桌上显示神通。“概率论产生于赌博”，这个观点是错误的或者说是不完全对的。“赌博问题”和“理性思考”是概率论产生的两个必要条件，而后者更重要。犹如苹果落地千千万，而只有牛顿从中发现了万有引力定律。不少学者错误地认为，帕斯卡、费马和惠更斯三人一起讨论了概率问题，而后者仅是将前二者的结果著书立说。从该书的撰写过程来看，惠更斯几乎全是自己独立解决这些概率问题的，虽帕斯卡、费马间接给他提供了一些问题，但均无解答过程。概率史界认为，帕斯卡与费马的通信标志着概率论的诞生。然而他们的通信直至1679 年才完全公布于世，故惠更斯的《论赌博中的计算》标志着概率论的诞生。因此，不少学者宣称惠更斯为概率论的正式创始人。惠更斯的《论赌博中的计算》不仅是第一部概率论著作，而且是第一个把该学科建立在公理、命题和问题上而构成一个较完整的理论体系，第一次对以前概率论知识系统化、公式化和一般化。该书为概率论的进一步发展奠定了坚实的基础。

1657 年9 月《论赌博中的计算》出版后立即得到学术界的认可和重视。该书在欧洲多次再版，作为概率论的标准教材长达50 年之久。直至1713 年雅各布·伯努利的《猜度术》出版才遏制住该书的再版，然而该书的影响还在继续。因《猜度术》的第一卷就是《论赌博的计算》的注释，并借此建立了第一个大数定理。法国数学家棣莫弗的《机会论》也是在该书的基础上，由二项分布的逼近得到了正态分布的密度函数表达式。拉普拉斯在此基础上给出古典概率的定义。因此，惠更斯的概率思想对古典概率的影响是重要而持久的，其方法可以看作那一时期的特点。但是，至于什么是“理想理论”，需要考虑它的历史发展阶段，不能苛求古人，也不能执于一偏。

尽管惠更斯的《论赌博中的计算》已出版300 余年了，但其科学的思想方法已跨越时空在数学教育尤其是概率论的学习中散发着无穷的力量。了解其内容有助于我们学习和应用概率论这一重要的数学分支。正如拉普拉斯所说“一门开始于研究赌博机会的科学，居然成了人类知识中最重要的学科，这无疑是令人惊讶的事情”。

惠更斯的理论是很完善的，他以机会问题为研究对象，以期望作为基本概念和表达工具，建立了概率(期望)的运算定理，他总结了前人的代数和组合方法，还引入了方程新方法，将具体赌博问题的分析提升到一般化的高度，这是有史以来第一个具有严格科学体系的随机理论成果。

到17 世纪，不少学者已对赌博中的某些问题进行了讨论，并挖掘了其中的数学原理。但对当时的大多数数学家来说，概率论是庸俗的赌博游戏，难登大雅之堂。正是社会的发展及其需要，才推动了概率论的发展。如果没有社会的需要，概率论至今恐怕仍然只能在牌桌上显示神通。“概率论产生于赌博”这个观点是错误的或者说是不完全对的。“赌博问题”和“理性思考”是概率论产生的两个必要条件，而后者更重要。犹如苹果落地千千万，而只有牛顿从中发现了万有引力定律。

不少学者错误地认为，帕斯卡、费马和惠更斯三人一起讨论了概率问题，而惠更斯仅是将前二者的结果著书立说。从该书的撰写过程来看，惠更斯几乎全是自己独立地解决了

这些概率问题，虽然帕斯卡、费马间接地给他提供了一些问题，但均无解答过程。

惠更斯的概率论著作出版后受到同时代数学家的欢迎。在雅各布·伯努利的《猜度术》问世之前，一直无人能超过惠更斯的概率理论，它一直是概率论的最佳读本。

就帕斯卡与费马的天才而言，他们若对概率问题进行彻底的研究，则无疑能够为经典概率建立起一般性的理论。帕斯卡当时糟糕的精神状态使他无法继续对数学的研究，他对费马提出深入讨论概率的建议，反应也非常冷淡。费马后来由于对数论产生了兴趣，也就放弃了对概率问题的思考。从时间上讲，他们两人对概率的研究要早于惠更斯，不过在深度方面则有所不及。

1.5 雅各布·伯努利的《猜度术》研究

伯努利家族，又译贝努利家族。17~18 世纪瑞士巴塞尔的数学和自然科学家的大家族，祖孙三代，出过十多位数学家。原籍比利时安特卫普，1583 年遭受天主教迫害，迁往德国法兰克福，最后定居巴塞尔，主要成员的世系如图 1-2 所示。

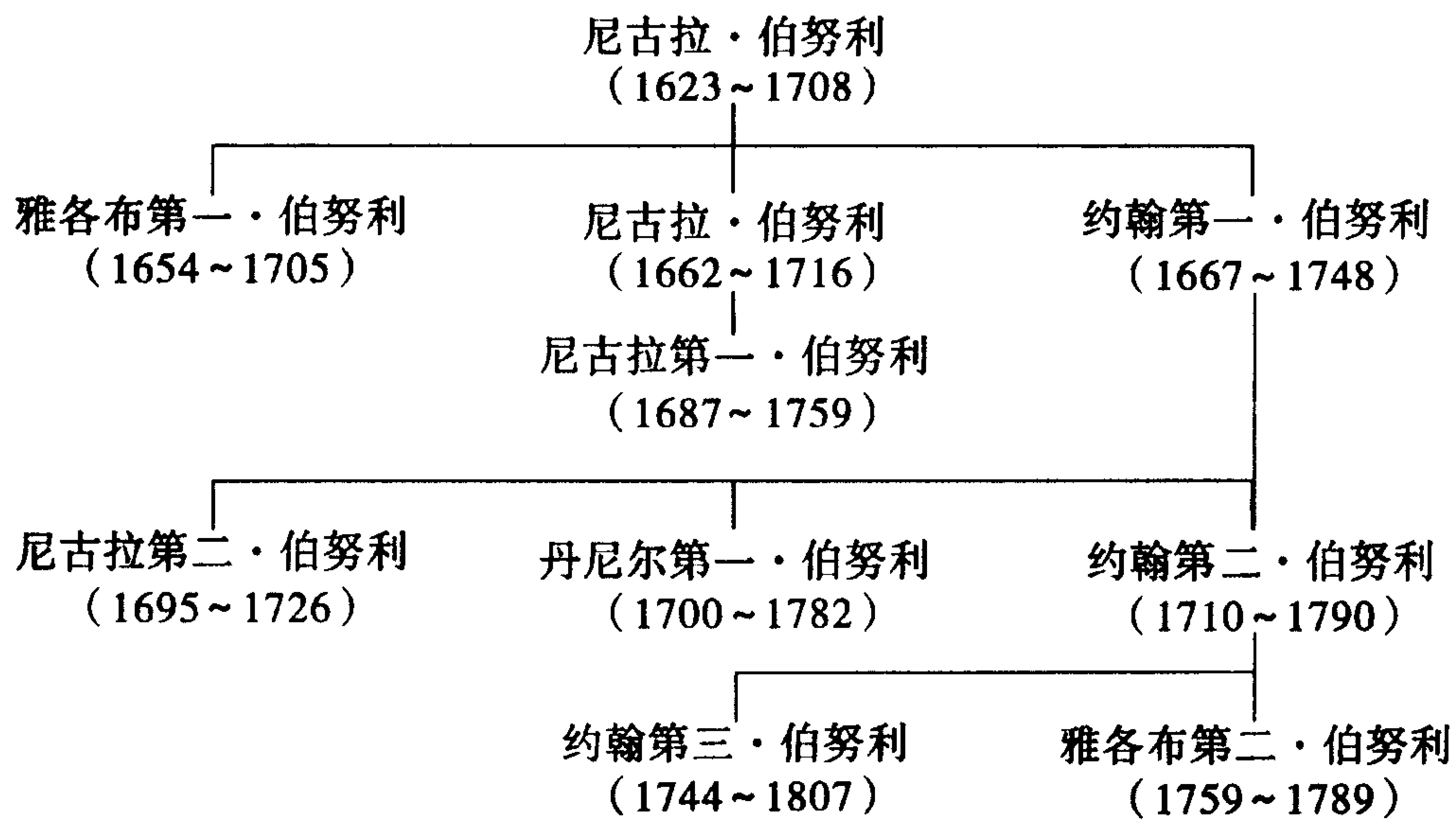


图 1-2

最重要的是雅各布第一·伯努利、约翰第一·伯努利和丹尼尔第一·伯努利。

雅各布·伯努利(Jacob Bernoulli，1654~1705)1654 年 12 月 27 日生于瑞士巴塞尔，1705 年 8 月 16 日卒于同地。最初遵从父亲的意见学神学，当他读了 R.笛卡尔、J.沃利斯的书后，颇受启发，兴趣转向数学。1676 年到荷兰、英国等处，结识当地学者，从 1687 年起直到去世，任巴塞尔大学教授。他和弟弟约翰第一·伯努利是 G.W.莱布尼茨的朋友，他们迅速掌握了莱布尼茨的微积分并加以发扬光大。雅各布在《学艺》上发表一系列的论文，1694 年他首次给出直角坐标和极坐标下的曲率半径公式，这也是系统地使用极坐标的开始，1690 年他提出悬链线问题，后来又改变条件，解决了更复杂的悬链问题。1694 年的论文讨论了双扭线的性质，伯努利双扭线因此得名。1695 年他提出著名的伯努利方程。

雅各布第一对对数螺线深有研究，他发现对数螺线经过各种变换后，结果还是对数螺

线。在惊叹这曲线的奇妙之余，遗言要将这曲线刻在墓碑上，并附以颂词："纵使变化，依然故我"。雅各布的巨著《猜度术》（1713）的出版，是组合数学及概率论史的一件大事，书中给出的伯努利数有很多应用。还有伯努利定理，这是大数定理的最早形式。

约翰第一·伯努利1667年8月6日生于巴塞尔，1748年1月1日卒于同地。最初学医，同时研习数学。1691年到巴黎，曾为洛必达的私人教师。现今求不定式极限的洛必达法则，实出自约翰。1705年接替其巴雅各布任巴塞尔大学教授。1691年解出悬链线问题。1696年，他向全欧洲数学家挑战，提出最速降曲线问题："一质点受地心引力的作用，自较高点下滑至较低点，不计摩擦，问沿着什么曲线，时间最短？"问题的难度在于和普通的极大极小值求法不同，它是要求出一个未知函数（曲线）来满足所给条件。洛必达、莱布尼茨、I.牛顿、雅各布第一·伯努利都给出这个问题的解答，后来引起变分法的产生。

雅各布第一、约翰第一兄弟两人写出了数学史中最重要的兄弟成功的故事，但两人的关系并不和谐。恰恰相反，在数学中，他们两人中每一个人都是另一个人强劲的竞争对手，两人为了胜出对方一筹而斗力，甚至到了可笑的地步。两人经常在相同的领域工作，并经常相互争论。这些争论促进了数学的发展。但由于双方过分敏感自尊，性格暴躁，相互批评指责又过于尖刻，使兄弟之间时常造成不快。

尼古拉第二·伯努利，是约翰第一·伯努利的儿子，13岁入巴塞尔大学，1715年取得法学硕士学位。1725年同其弟弟丹尼尔第一·伯努利一起应邀到圣彼得堡去。他到圣彼得堡后曾提出一个概率论问题，后来以圣彼得堡问题著称，可惜次年就死在那里。

丹尼尔第一·伯努利，1700年2月8日生于荷兰格罗宁根，1782年3月17日卒于巴塞尔。丹尼尔25岁就成为圣彼得堡科学院数学教授，他最早的论著是解决黎卡提方程（1724）。他在概率论、偏微分方程、物理等方面均有贡献。曾获法国科学院奖金10次之多。他的《流体动力学》1738年出版，这是作为流体动力学基础的"伯努利定理"的出处。1733年他回到巴塞尔，教授解剖学、植物学和自然哲学。

雅各布·伯努利一生最有创造力的著作就是1713年出版的《猜度术》，在这部著作中，他提出了概率论中的"伯努利定理"，该定理是"大数定律"的最早形式，由于"大数定律"的极端重要性，1913年12月圣彼得堡科学院曾举行庆祝大会，纪念"大数定律"诞生200周年。

从雅各布·伯努利与莱布尼兹的通信中，可知他写《猜度术》这一著作是在他生命的最后两年。在1705年他去世时，此书尚未整理定稿。由于伯努利兄弟在科学问题上的过于激烈的争论，致使双方的家庭也被卷入，以至于雅各布死后，整理和出版遗稿的工作，迟迟未能实现。他的《猜度术》手稿被他的遗孀和儿子在外藏匿多年，不愿把整理出版的事委托给约翰，后来又拒绝了欧洲一位富有的学者捐资出版的建议。最后在莱布尼兹的敦促下，雅各布的儿子才于1712年10月开始整理并印刷《猜度术》。而此时，雅各布的侄子尼古拉斯（N.Bernoulli I，1687~1759）正在法国帮助蒙特摩（P.R.Montmort，1687~1719）准备《赌博游戏的分析随笔》第二版。1713年5月，当尼古拉斯回到巴塞尔时，《猜度术》的整理和印刷工作已接近尾声。直到此时，雅各布的儿子才敢与他这位大堂兄打交道，请他帮助韵饰定版，此时所作任何补充都太晚了，为了使该书尽早付梓，尼古拉斯只匆匆写了一篇两页的序言，并为较严重的印刷错误编了一张勘误表。1713年8月，在雅各布死后8年，才得以出版，几乎使这部经典著作的价值受到损害。而不少概率论史家据此就认为是尼古拉斯整理出版了《猜度术》，如著名的哲学家和数学教授沃尔夫（C.WOLff）在其《数学辞典》（Leipzig,

1716）中就持有这一观点，同样的错误也发生在托德亨特（Todhunter）等人的论述中。

雅各布·伯努利的遗著《猜度术》奠定了概率论作为一门独立数学分支的基础，其中首次提出了后来称为“伯努利定理”的极限定理，刻画了大量经验观测中呈现的稳定性，伯努利定理作为大数定律的最早形式在概率论发展史上占有重要地位。因而雅各布·伯努利的著作《猜度术》是概率论发展史上的具有“里程碑”式的事件。

1.《猜度术》内容简介

《猜度术》除前言共有 306 页，呈小四开本形式，内容分成四个部分。在第一部分(1~71 页)中对惠更斯的著作《论赌博中的计算》作了仔细的注解，比惠更斯的原文长四倍。第二部分是关于排列组合的比较系统的论述(72~137 页)。第三部分利用前面的知识，讨论了一些使用骰子等的博奕问题(138~209 页)，第四部分(210~239 页)涉及概率论在社会、道德和经济等领域中的应用，其中包含了本书的精华——以他的名字命名的大数定律，可以说，本书对后世的影响如此之大，原因全在这一只含 30 页的部分。要是没有它，本书也可能像一些概率论早期著作一样湮没无闻，或至多当作一个一般性的事件来看待。本书附录，一部分是关于无穷级数的五篇论文，还包含一个 36 页(271~306 页)以与友人通信的形式讨论网球比赛中的计分问题。

在第一部分(1~71 页)中，雅各布对惠更斯的《论赌博中的计算》一书做了详尽的评注，指出了前人在概率推理时暗含的一些假定，如，有放回抽取、重复赌博中的独立性、加法原则等。他给出了著名的伯努利试验及多重伯努利试验中成功次数的概率分布，即二项分布。这部分内容比惠更斯的原文长四倍还多，这使人们很容易相信雅各布写此书在很大程度上是受惠更斯的影响。

他的评注被认为比惠更斯的原文更有价值。如：注文中给出如下结果：

(1)一次掷 n 个骰子得 m 点，共有方式个数等于 $(x+x^2+x^3+x^4+x^5+x^6)^n$ 展开式中 x^m 的系数。

(2)设一次试验中成功与失败的概率分别是 $\frac{b}{a}$ 和 $\frac{c}{a}$，则 n 次试验中成功 m 次的概率是 $(\frac{b}{a}+\frac{c}{a})^n$ 的展开式中前 $n-m+1$ 项之和

$$(\frac{b}{a})^n+C_n^1(\frac{b}{a})^{n-1}(\frac{c}{a})+C_n^2(\frac{b}{a})^{n-2}(\frac{c}{a})^2+\cdots+C_n^{n-m}(\frac{b}{a})^m(\frac{c}{a})^{n-m}.$$

(3)惠更斯命题 14 中问题的解：设想有无穷多人参加掷骰子。序数为奇数者掷得 6 点，则赢；序数为偶数者掷得 7 点者，则赢。一对骰子掷一次的所有 $a=36$ 种可能情形中，6 点的情形共有 $b=5$ 种，不是 6 点的共有 $c=31$ 种，不是 7 点的情形共有 $f=30$ 种。于是各投掷者的期望值如表 1-6 所示。

表 1-6

Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	Ⅶ	Ⅷ	…
$\frac{b}{a}$	$\frac{ce}{a^2}$	$\frac{bcf}{a^3}$	$\frac{c^2ef}{a^4}$	$\frac{bc^2f^2}{a^5}$	$\frac{c^3ef^2}{a^6}$	$\frac{bc^3ef^3}{a^7}$	$\frac{c^4ef^3}{a^8}$	…

现设序数为奇数者为同一人甲，序数为偶数者为同一人乙，于是甲、乙期望值分别为两个几何级数之和：

$$\frac{b}{a}+\frac{bcf}{a^3}+\frac{bc^2f^2}{a^5}+\cdots=\frac{ab}{a^2-cf}=30$$

$$\frac{ce}{a^2}+\frac{c^2ef}{a^4}+\frac{c^3ef^2}{a^6}+\cdots=\frac{ce}{a^2-cf}=31$$

(4)一般性解决了惠更斯所提第 5 个问题："甲、乙各有 12 个筹码，掷三个骰子。如果掷得 11 点，则甲必须给乙 1 个筹码；如果掷得 14 点，则乙必须个甲 1 个筹码。最先得到所有筹码者为胜者。证明：甲、乙获胜的概率之比为 244 140 625：282 429 536 481" 。

设甲、乙拥有筹码数分别为 m 和 n 。如果记甲拥有 x 个筹码时最终赢得对方所有筹码的概率为 u_x ，则有递推关系

$$u_x=\frac{a}{a+b}u_{x+1}+\frac{b}{a+b}u_{x-1},$$

注意到 $u_0=0$ ， $u_{m+n}=1$ ，不难得到

$$u_x=\frac{a^{m+n}-a^{m+n-x}b^x}{a^{m+n}-b^{m+n}},$$

于是甲最终赢得全部筹码的概率为

$$u_m=\frac{a^n(a^m-b^m)}{a^{m+n}-b^{m+n}},$$

同理，乙最终赢得全部筹码的概率为

$$u_n=\frac{b^m(a^n-b^n)}{a^{m+n}-b^{m+n}}.$$

在第二部分中，雅各布对前人有关组合分析的工作进行了系统性的整理与解释，并给出了著名的伯努利数(自然数幂和)的求法，给出了点数问题的解，与帕斯卡在《论算术三角形》中的结果相一致。这部分内容最终成为 18 世纪有关组合分析理论方面的最流行的教科书。

在第三部分中，雅各布将惠更斯讨论的赌博问题作了更进一步的扩展。他归纳出 24 个典型的赌博问题，并利用加法与乘法公式、组合方法、条件期望、递归方法对这些问题进行了解答。如第 14 个问题："甲掷骰子，第一次掷得几点，接着就重复掷几次。如果掷得点数之和超过 12，则甲赢得全部赌金；如果点数之和小于 12，则甲得不到赌金。求甲的期望值。"伯努利求得答案为

15 295/31 104

从前三部分的内容来看，雅各布不愧为伟大的数学家，他将前人的成果进行了系统性的整理、解释与扩展，并使《猜度术》成为截止到 18 世纪初人类有关概率论知识的最全面的总结。然而使《猜度术》成为概率统计发展史上的里程碑却是仅有 30 页的第四部分。

在第四部分中，雅各布对概率论的发展做出两项革命性的贡献：

(1)第一次对主观概率的概念进行了清晰的解释。

(2)第一次给出了大数定律的证明。

下面我们分别对这两个主题进行较为详细的介绍。

像同时代的大多数科学家一样，雅各布是形而上学决定论的坚定支持者。他认为“全能全知的上帝已将所有的事情全部安排好了，它没有留下任何随机现象，进而所谓的随机现象无非是我们无知的体现”。他认为概率是确定程度的度量，而确定程度是依赖一个人的知识的。他对此进行了举例说明。“当骰子的位置、速度、离桌子的距离以及离开投掷者手时的动量给定时，骰子的结果就应是投掷的结果，而不会是其他结果；同样，当质量、位置、动量、方向以及风、蒸汽、云的速度给定时，并且还知道他们相互影响的机械定律，那么明天的暴风雨只能以它出现的方式出现。与‘食’(eclipses)的现象由发光天体的运动所决定一样，这两种现象一点也不缺少必然性。然而经验却只主张对‘食’按照必然性来预测，而对骰子的结果及明天的暴风雨按照随机性来预测。但这个事实只不过说明我们对于决定后一类现象的数据(特别是关于自然的)还没有充分的获知而已。”……“因此一个人可能在某一时刻认为一个现象是随机的；而在另外一个时刻，当这个现象发生的原因搞清楚后，这个人则会认为这个现象是必然的。所以随机性特别依赖于我们的知识”。由此我们可以看出雅各布的概率的概念完全是主观(subjective)的。然而也有人认为雅各布是坚定的频率学派学者。实际上，几乎每个学派都宣称雅各布属于自己的学派。对此 Hacking 的解释是“像我们大家一样，他被所有这些不一致的思想深深地吸引住了，以至于他自己也不知道到底应持何种观点。”

在给出了概率的定义之后，雅各布又引进了一个对后世统计推断思想影响巨大的概念，即所谓的“有把握确定”(moralcertainty)。他写道“确定就是概率接近于完全确定，所以有把握确定的事件是不能认为不发生的事件”。如果一件事情是有把握确定的，概率为 999/1000，那么有把握不可能(moralimpossible)的概率就只有 1/1000 了”。显然雅各布“有把握确定”就是今天数理统计中的“置信水平”。“有把握不可能事件”就是“小概率事件”。

由于持有主观概率的思想，雅各布认为概率就是从已知的“证据”(Proof)或“前提”对“结论”确定程度的度量。因此他花了大量的篇幅(第四部分的第三、四章)讨论了如何区分证据和利用证据进行科学的推断。在给出个有关利用证据去判断凶手的例子之后，雅各布写到“从前面的讨论可以看出，证据的力度依赖于多种情况(case)，每种情况都可能存在也可能不存在；可能表明某件事也可能不表明这件事；甚至可能表明相反的事。所以，证据所产生的确定程度或概率可以利用这些情况(case)按照第一部分讨论的方法来计算，就像赌博游戏中习惯使用的方法一样”。

雅各布的这些讨论被认为促使了拉普拉斯“不充分推理原”(principle of insufficientr eason)的产生；同时也使雅各布被认为是由 Keynes、Carnap 创建的归纳逻辑学派的先驱。

2.概率的观点

伯努利在书中对概率的概念花了不少篇幅进行讨论，发表了一些对以后有影响的观点。他也是采取把概率区分为客观概率和主观概率的立场。有些论述也兼收了前人的观点，值得注意的新看法，归纳起来有以下几点。

(1)在客观概率中，他明确区分了“可以先验地计算”的概率和“后验地计算”的概率。前者是指基于对称性即等可能性的考虑去计算，它不用进行实际观察，如掷一个均匀骰子，则基于对称性，“掷出偶数点”的概率为 1/2。这不待多次掷骰子去观察其频率即可推知。而“出生男孩”这一事件的概率，则须观察大量的新生婴儿才能算出。用现代语言说，伯

努利在客观概率中，明确区分了现代称之为“古典概率”和“统计概率”这两种情况。注意此处“先验”一词与 Bayes 统计中的“先验分布”无关。

(2)伯努利对事物采取了一种机械决定论的立场，或是说，世界上的事物受严格的因果律的支配。他分析掷骰子这个例子，认为：若把一切有关条件全弄准了，则投掷结果也唯一确定了。因而并不存在随机性，后来拉普拉斯也是采取这一立场，反驳这一说法的一个可能的论点是：根据量子力学中海森堡的“测不准原理”，你不可能把一切有关的初始条件同时弄准确，因而也就失掉了精确预言骰子投掷结果的根据。然而，爱因斯坦对测不准原理有怀疑。他曾说过“上帝不掷骰子”。

(3)伯努利引进了所谓“道德确定性”(moralcerfainty)的概念。一件事情，虽不能断言其绝对的确定性，但若它被公认为以极大的可能性以至几乎不可能不发生，就称为有“道德确定性”。这个概念日后在统计推断中有重大影响，如$(1-\alpha)$置信区间。现在我们把这说成“事实上的确定性”(practicalcertainty)，或者“小概率事件原理”。

(4)在涉及主观概率的问题中，当有若干种可能性存在，而没有任何论据去支持给予某一特定的可能性以优先的考虑时，伯努利提出给予这些可能性以同等的概率。这一思想后来在贝叶斯(均匀先验分布)和拉普拉斯(不充分理由的原则)中得到发展。在 Bayes 学派统计学的发展中起了重大的作用。

不过有一点需要指出的是，雅各布的 300 年的这部著作对于现代统计学的研究依然具有指导作用。从他认为概率论(统计学)可以应用到所有经验科学领域和人们每天的日常生活中的思想，我们不难发现这样的经验事实：统计学是人类认识不确定现象的工具以及行为决策的基础。除了统计学家，大多数人也都或多或少地从生活或科学实践与社会实践中掌握一些统计知识(如，甚至最笨的人也会理解大数定律；又如，2000 年前人类已经知道用算术平均数调整数据了，而统计学的诞生只有 300 年左右)，不过，这种知识可能是非理性的、经验的、直觉的，当然也可能是现行统计理论未包含的。所以统计学应重视对生活和实践的学习。

3.伯努利大数定律

雅各布·伯努利在《猜度术》中关于大数定律的论述如下：

“我们现在已达成这样的共识，好像无论对任何事件做出一个准确的推测，唯一必要的是先精确的计算事件的可能结果数目，然后确定其中某一结果比另一结果更可能发生的可能性有多大。不过我们所面临的问题也随之产生，因为这种方法仅适用于极罕见的现象。实际上，这种方法几乎是那些与机会游戏相联系的现象所专有的，机会游戏的最初发明者通过固定导致输和赢的结果的数目，让这些结果事先已知，并安排游戏使每一结果发生都是等可能的方式，设计这些游戏，使得所有玩家有相同的赢的机会。至于别的大量的属于由自然规律或人为意志所支配的现象，则决不是这种现象。例如，在掷骰子的游戏中，由于掷出各种点数的可能性与骰子的面一样多，这种游戏的可能结果（或掷出的各种点数）都已知。此外，但骰子的每一个面都具有相同的构造，而骰子的密度是均匀分布时，所有可能结果的发生都是等可能的。（这种情况下，没有理由要求骰子的一个面比另一个更容易发生，但是如果骰子的各个面具有不同的构造，且骰子的一部分是用比其余部分更重要的材料做成的，则又另当别论）。类似地从一个可能结果数目已知的坛子里取出白球或黑球，能够断言坛子里任何一个球都是等可能的被取出：因为知道坛子里每一类球有多少个，

所以没有理由认为坛子里这个或那个球应该比别的球更容易被取出来。然而，我要问，人能够查出不同年龄折磨人体的许多疾病，视它们为所有可能的结果，能说出某一种疾病比另一种疾病更可能致命的可能性有多大吗?比如瘟疫比浮肿，或者浮肿比发烧，并在此基础上，关于未来一代人生与死的关系做出预言吗?或者，谁能列举出每天大气层经历的不可数的变化，并由此预言今天，乃至未来一个月，或一年内天气情况?再者，谁能自称可以深刻的透视人的思维或人体的奇异结构，使得在游戏中完全或部分地依靠玩家的思维的敏捷性或身体的灵活性，大胆预言这个或那个玩家会赢或会输?这些或类似的预言依赖于完全不分明的因素。这些因素通过它们之间相互关系的无尽的复杂性，不断的欺骗我们的辨别力，使得试图沿着这条路线进行下去简直是无意义的。

不过，存在另一条途径，将引导我们去寻找所期待的结果。在不能确定一个‘先验’时，至少使我们能确定一个‘后验’。即从大量已观测到的例证的结果来确定这个‘后验’，为做到这一点，必须假定在相同的条件下，一个事件发生（或不发生）遵从过去同类事件被观测到的相同规律。例如，我们观测具有相同体质和相同年龄 300 人的生与死，结果在 10 年内有 200 人死亡，而其余的人还活着。这时，我有确切的理由断言，有三分之二这样的人在十年内辞世的可能性相同，而其余三分之一的人则有机会生存长于 10 年。类似的，如果任何一个人 观察了一年内各时期天气的变化情况，并记录下有多少晴天，多少雨天；或者反复观看两个玩家的局况并留意两个玩家各自赢了多少次，于是，仅基于这些观测，在于过去相同条件的假定下，他就能确定相同结果在将来发生与否的比例。

这种通过观察来确定结果数目的经验方法既不是新的，也不是不常见的。在第十二章，仿效 L'art de penser 那位聪明而有天资的作者的，描述了一个类似的方法。在日常生活中，我们都能看到相同的原理在起作用。每个人都清楚，利用任何单个观测作为基础来预言一些（未来）事件是不充分的，而应要求有大量的观测。一些例证表明，一个没有受过训练的人，凭天生的直觉，也会清楚地知道，可利用的有关观测的次数越多，发生错误的风险就越小。尽管我们都以非常自然的方式意识到这一点。但是，这个原理的科学证明却一点也不简单。因此，在此指出这一点是我义不容辞的责任。如果我仅限于证明这一任何人都熟悉的原理，我深感自己做得太少了。有些问题更需要考虑，而对人们来说，这些问题也许还没有想到。‘所要探讨的是，是否随着探讨次数的增大，记录下来的赞成与不赞成例数的比值接近真实比值概率也随之不断增加，使得这个概率最终将超过任意确信度；’或者是否问题好像有一条渐近线。这将意味着存在一个特殊的确信度，通过任意增加观测次数，上述概率也不能超过这个确信度。譬如，我们不能确信已确定的结果的真实比值大 $\frac{1}{2}$，$\frac{2}{3}$ 或 $\frac{3}{4}$。以下的解释将使我们所说的含义更清楚：我们有一个罐子，里面装有 3 000 个小白卵石，和 2 000 个小黑卵石，我们希望从经验上确定罐中白卵石与黑卵石的比例——我们不知道的事情，方法是将卵石一个接一个的取出，记录下来取出了多少白卵石和多少黑卵石。（我提醒你们，这个取石过程的一个重要要求是，每次取出一个卵石，并记下其颜色，再取下个卵石之前，把这个卵石再放回罐中，使得罐中卵石的数量保持常数）。现在，我们问，是否可能通过无限扩大试验，比如做 10 次，100 次，1 000 次等，取出的白卵石数目与黑卵石数目之比？（最终是“必然”）和罐中白卵石和黑卵石数目的真实比例取相

同值（3：2），还是取出的白卵石数目与黑卵石数目之比将取不同值？如果回答是'不'，则我承认，试图通过观测来确定每个结果的例证数目（即白卵石与黑卵石的数目）可能要失败。通过这种办法，我们最终能达到内心确信的必然性……这时，我们几乎能很精确的确定例证的数目，好像我们事先知道这些数目似的。公理 9（在前一章提出的）表明，在我们的日常生活中，内心确信的必然性被看作是绝对的确定性，这种考虑能使我们对任何带有偶然性的事件可以做出预测，其科学程度不比对机会游戏所做出的预测差。例如，如果我们用大气层或人体来代替上面说的罐子，在其内部隐藏着大量的各种各样的变化过程或疾病，就如同罐子里隐藏着卵石一样，这时，对这种现象，我们仍能够通过观测来确定其中某事件将比另一事件更频繁发生的可能性有多大。

应该指出，放映不同结果数目之间世纪关系的比例——我们正试图通过观测来确定的比例——绝不可能精确的获得。这一点恐怕不完全能被理解。因为如果精确获得是可能的话，那么控制原理将与我所断言的相对立：即观测作的越多，我们发现正确不利的可能性越小。我们获得的比例仅是近似：他必须有两个限定来定义，如果我们取两个比例，$\frac{301}{200}$和$\frac{299}{200}$，$\frac{3001}{2000}$和$\frac{2999}{2000}$，或是任意两个类似的比例，其中一个稍小于$1\frac{1}{2}$，而另一个，稍大于$1\frac{1}{2}$，显然，通过大量重复观测所得到的比例降落在$1\frac{1}{2}$这个比值的这些限之间而不是这些限之外，且这些限能达到我们所希望的可能程度。

这正是经过 20 年的深思熟虑之后，我决定在此发表的问题。如果从现在直至永远，所有的事件都被连续的观测到（靠这点，可能性终将变成必然性），将会发现世界上每件事情发生都有着明确的原因和遵循明确的法则，甚至对看来相当偶然的事件，我们被强迫假定一定的必然性，似乎有点像命中注定似的。据我所知，这正是柏拉图当时所断言的，依然普遍循环的教条，在经历无数的世纪后，每一事物将返回他最初的状态。"

在伯努利《猜度术》的第四部分以大数定律而引人注目，首次提出了后来以"伯努利定理"命名而著称的极限定理，刻画了大量经验观测中呈现的稳定性，伯努利定理作为大数定律的最早形式在概率论发展史上占有重要的地位。

缸中有a个白球，b个黑球，摸到白球的概率$p=\frac{a}{a+b}$。有放回地从缸中抽球N次，记录抽到白球的次数为X，以$\frac{X}{N}$去估计p。这个估计法是现今数理统计学中最基本的方法之一。

伯努利企图证明的是：用$\frac{X}{N}$估计p可以达到事实上的确定性——他称为道德确定性。其确切含义是：任意给定两个数$\varepsilon>0$和$\eta>0$，总可以取足够大的抽取次数N，使事件$(|\frac{X}{N}-p|>\varepsilon)$的概率不超过$\eta$。这意思很显然：$|\frac{X}{N}-p|>\varepsilon$表明估计误差未达到制定的接近程度$\varepsilon$，但这种情况发生的可能性可以随心所欲地小(代价是加大N)。为忠实于伯努利的表达形式，应指出两点：一是伯努利把ε限定为$(a+b)^{-1}$，虽然其证明对一般ε也有效。他作这一限定与所用缸子模型的特殊性有关：必要时把缸中的白球、黑球分别改为ra和rb

个，则 p 不改变，$(a+b)^{-1}$ 改为 $\frac{1}{ra+rb}$，只须取 r 足够大，可使此数任意小。其次，伯努利要证明的是：对任给 $c>0$，只须抽取次数 N 足够大，可使

$$P(|\frac{X}{N}-p|\le\varepsilon)>cP(|\frac{X}{N}-p|>\varepsilon)$$

这与前面所说是一回事。

因为由上式得

$$P(|\frac{X}{N}-p|>\varepsilon)<(c+1)^{-1}$$

取 c 足够大可使它小于 η。

伯努利上述对事实上确定性的数学理解，有一个很值得赞赏之点，即他在概率论的发展刚起步的阶段，就给出了问题的一个适当的提法。因为，既然我们想要证明的是当 N 充分大时，$\frac{X}{N}$ 可以任意接近，则一个看来更直截了当的提法是

$$\lim_{N\to+\infty}\frac{X}{N}=p\text{ ,}$$

而这不可能实现。因为原则上不能排除“每次抽到白球”的可能性，这时 $\frac{X}{N}$ 总为 1，不能收敛于 $p<1$。或者退一步：要求 $\lim_{N\to+\infty}\frac{X}{N}=p$ 式成立的概率为 1，这个结论是对的，但直到 1909 年才由波莱尔证明，其难度比伯努利的提法大得多。设想如果当是伯努利就采用这个提法，他也许不一定能在他有生之年完成这一工作。布莱尔的结论比伯努利的强，故现在我们分别称为强大数律和弱大数律。

如今具有概率论初步知识的人都知道，伯努利大数律是切比雪夫不等式的简单推论。但在伯努利时代尚无方差概念，更不用说这一不等式了。伯努利采用的是直接估计概率的方法，大意如下：

令

$$A_0=P(Np<X<Np+N\varepsilon)\text{ ,}$$
$$A_k=P(Np+kN\varepsilon<X\le Np+(k+1)N\varepsilon)\text{ ,}\quad k=1,2,\cdots\text{ ,}$$

秩序证明：当 N 充分大时有

$$A_0>c(A_1+A_2+\cdots)\text{ ,}$$

这就解决了 $X<Np$ 的边，对 $X<Np$ 的一边如法炮制，就可得到伯努利大数律。

在第四部分的最后，雅各布开始讨论他反复思考了近 20 年的极限定理。在伯努利时代人们只能利用等可能假设计算赌博游戏中的概率，而在其他领域，如生命统计、保险，人们通常只能利用相对频率来估计概率了。雅各布意识到等可能的条件无法在赌博之外的领域得到验证，因此必须建立一种新的可在赌博领域之外计算概率的方法。

在考察同一事件在相同条件下观测到的相对频率时，雅各布表述了这样的经验事实：“有时即使是一个最笨的人也会根据本能，在没有任何事前教育的情况下(这确实令人吃惊)，坚信观测的次数越多，偏离目标的危险就越小”。雅各布意识到如果能对这个事实给

出理论上的证明，那么我们就可以放心地在任何领域利用相对频率来计算概率了。接下来雅各布决定对这个大家都知道，但都未意识到的事实进行证明。

在证明中，雅各布将上面提到的事实转化成一个特殊的例子。

考虑了一个随机试验，它有

$$t = r + s$$

种等可能的结果，其中有利结果为 r 种，这样有利的概率就是

$$p = \frac{r}{r+s} = \frac{r}{t}$$

将这个随机试验独立重复地进行 n 次，设 X_n 为有利结果出现的频率，雅各布证明了如下结论：

对任意给定的正实数 c，只要 $n = kt$ 充分大，或 $k \geq \max\{k(r,s,c), k(s,r,c)\}$，$k(r,s,c)$ 为满足条件 $k(r,s,c) \geq \frac{m(r+s+1)-s}{r+1}$ 的最小正整数，m 为满足条件 $m \geq \frac{\ln[c(s-1)]}{\ln[(r+1)/r]}$ 的最小正整数，就有

$$P\{|X_n - p| \leq \frac{1}{t}\} > \frac{c}{c+1}$$

关于雅各布的证明，需要指出的是，雅各布的定理与现代的大数定律有两点不同：

(1)为了证明的方便，雅各布假定 n 是 t 的整数倍。不过这一点并不会对结论产生实质性的影响。

(2)结论中频率与概率的误差不是给定的任意小的正数 ε，而是 $1/t$。

但 Stigler 解释这可以通过对 r, s 放大相同的倍数的方法来解决 $1/t$ 任意小的问题。

雅各布证明非常具有古典概率的味道，虽然比现代的证明长一些，但看上去更自然。

在结束上面的证明之后，雅各布给出了一个简单的例子。

设 r =30，s =20，则某事件出现的概率为 30/50，计算该事件出现次数与全部次数之比介于 29/50 和 31/50 之间的概率。其结果为：

(1)若 c =1 000，那么 n 至少等于 25 550，概率值为 0.999。

(2)若 c =10 000，那么 n 至少等于 31 258，概率值为 0.9 999。

(3)若 c =100 000，那么 n 至少等于 36 966，概率值为 0.99 999。

至此，《猜度术》全书结束。

4.历史评价

《猜度术》是雅各布·伯努利一生中最有创造力的著作，该书的出版标志着概率论已建立在稳固的数学基础上并成为一门独立的数学分支。对于《猜度术》史学家们已有各种评价，在此本人不想多谈，正如美国概率论史学家海金（Hacking）所说此书为“概率概念漫长形成过程的终结与数学概率论的开端。”时至今日，雅各布的概率思想在概率领域仍有着重要的影响。

不过有一点需要指出的是，雅各布的 300 年前的这部著作对于现代统计学的研究依然具有指导作用。从他认为概率论(统计学)可以应用到所有经验科学领域和人们每天的日常生活中的思想，我们不难发现这样的经验事实：统计学是人类认识不确定现象的工具以及行为决策的基础。除了统计学家，大多数人也都或多或少地从生活或科学实践与社会实践

中掌握一些统计知识(如，甚至最笨的人也会理解大数定律；又如，2000 年前人类已经知道用算术平均数调整数据了，而统计学的诞生只有 300 年左右)，不过，这种知识可能是非理性的、经验的、直觉的，当然也可能是现行统计理论未包含的，所以统计学应重视对生活和实践的学习。

1.6 棣莫弗的《机遇论》研究

棣莫弗(A.De Moiver，1667~1754)1667 年 5 月 26 日生于法国维特里的弗朗索瓦；1754 年 11 月 27 日卒于英国伦敦。

棣莫弗出生于法国的一个乡村医生之家，其父一生勤俭，以行医所得勉强维持家人温饱。棣莫弗自幼接受父亲的教育，稍大后进入当地一所天主教学校念书，这所学校宗教气氛不浓，学生们得以在一种轻松、自由的环境中学习，这对他的性格产生了重大影响。随后，他离开农村，进入色拉的一所清教徒学院继续求学，这里却戒律森严，令人窒息，学校要求学生宣誓效忠教会，棣莫弗拒绝服从，于是受到了严厉制裁，被罚背诵各种宗教教义。那时，学校不重视数学教育，但棣莫弗常常偷偷地学习数学。在早期所学的数学著作中，他最感兴趣的是 C.惠更斯关于赌博的著作，特别是惠更斯于 1657 年出版的《论赌博中的计算》一书，启发了他的灵感。

1684 年，棣莫弗来到巴黎，幸运地遇见了法国杰出的数学教育家、热心传播数学知识的 J.奥扎拉姆(Ozanam)。在奥扎拉姆的鼓励下，棣莫弗学习了欧几里得(Euclid)的《几何原本》(Elements)及其他数学家的一些重要数学著作。

1685 年，棣莫弗与许多信仰新教的教友一道，参加了震惊欧洲的宗教骚乱，在这场骚乱中，他与许多人一起被监禁起来。正是在这一年，保护加尔文教徒的南兹敕令被撤销。随后，包括棣莫弗在内的许多有才华的学者由法国移住英国。据教会的材料记载，棣莫弗一直被监禁至 1688 年才获释，并于当年移居伦敦。但据 20 世纪 60 年代发现的一份当时的材料，1686 年时棣莫弗已经到了英国。随后，棣莫弗一直生活在英国，他对数学的所有贡献全是在英国做出的。

抵达伦敦后，棣莫弗立刻发现了许多优秀的科学著作，于是如饥似渴地学习。一个偶然的机会，他读到 I.牛顿(Newton)刚刚出版的《自然哲学的数学原理》(Mathematical principles of natural philosophy)，深深地被这部著作吸引了。后来，他曾回忆起自己是如何学习牛顿的这部巨著的：他靠做家庭教师糊口，必须给许多家庭的孩子上课，因此时间很紧，于是就将这部巨著拆开，当他教完一家的孩子后去另一家的路上，赶紧阅读几页，不久便把这部书学完了。这样，棣莫弗很快就有了充实的学术基础，并开始进行学术研究。

1692 年，棣莫弗拜会了英国皇家学会秘书 E.哈雷(Halley)，哈雷将棣莫弗的第一篇数学论文“论牛顿的流数原理”(On Newton’s doctrine of fluxions)在英国皇家学会上宣读，引起了学术界的注意。1697 年，由于哈雷的努力，棣莫弗当选为英国皇家学会会员。

棣莫弗的天才及成就逐新受到了人们广泛的关注和尊重。哈雷将棣莫弗的重要著作《机遇论》(The doctrine of chances)呈送牛顿，牛顿对棣莫弗十分欣赏。据说，后来遇到学生向牛顿请教概率方面的问题时，他就说：“这样的问题应该去找棣莫弗，他对这些问题的研究比我深入得多”。1710 年，棣莫弗被委派参与英国皇家学会调查牛顿-莱布尼茨关于微积分优先权的委员会，可见他很受学术界的尊重。1735 年，棣莫弗被选为柏林科学院

院士。1754年，又被法国的巴黎科学院接纳为会员。

棣莫弗终生未婚。尽管他在学术研究方面颇有成就，但却贫困潦倒。自到英国伦敦直至晚年，他一直做数学方面的家庭教师。他不时撰写文章，还参与研究确定保险年金的实际问题，但获得的收入却极其微薄，只能勉强糊口。他经常抱怨说，周而复始从一家到另一家给孩子们讲课，单调乏味地奔波于雇主之间，纯粹是浪费时间。为此，他曾做了许多努力，试图改变自己的处境，但无济于事。

棣莫弗在87岁时患上了嗜眠症，每天睡觉长达20小时。当时有一个传说，棣莫弗的睡眠时间以每天1/4小时的速度递增，形成一等差级数。当24小时高睡不起时，他便在贫寒中离开了人世。

观察周围的现象会发现，大部分实际存在的随机变量都具有“中间大，两头小，左右对称”的特点。如测量某一长度的结果及其误差，某地区的年平均气温、平均雨量，在一定条件下生产的产品尺寸，弹落点与目标的距离，某农作物的产量，普通高考成绩，人的身高及智力水平等。这种随机变量所服从的分布称之为“正态分布”或“常态分布”，即“正常”情况下的随机变量总服从这种分布。不少书中还称之为“高斯分布”。若认为此分布的第一个发现者是数学王子高斯(C.F.Gauss，1777~1855)，那就张冠李戴了。因为早在高斯之前，正态概率曲线所呈现的规律即已为人所知，其发现者就是法国杰出的数学家亚伯拉罕·棣莫弗。

1.正态概率曲线产生的知识背景

棣莫弗是法国的加尔文教徒。1685年废除南特法令后，他迁居至伦敦，故其多半生是在英国度过的。尽管他只能靠做家庭教师来维持极贫困的生活，但他在学术研究上颇有成就。1697年，棣莫弗当选为英国皇家学会会员。1710年，被委派参与英国皇家学会委员会，调查牛顿(Isaac Newton，1642~1727)、莱布尼兹(Wilhelm Leibniz，1646~1716)微积分的优先权问题。1735年，当选为柏林科学院院士。1754年，他又被法国巴黎科学院接纳为会员。亚历山大教皇的诗作赞美了他。他一生最感兴趣的是概率论研究。最初是惠更斯的概率论著作，启发了他的灵感。之后，他又研究了蒙特摩(P.R.Montmort，1678~1719)的《随笔》。雅各布去世后，大数定律的精髓在学术界开始传播，由此棣莫弗对概率论兴趣倍增，并开始对这神秘的“机会”进行研究。1711年，他关于概率论的论文《关于游戏中机遇巧合的概率》以拉丁文连续发表在英国《哲学学报》1、2、3月号上。该论文含有26个问题及几个有关概率计算的注释。随后，棣莫弗将该文扩展成英文专著《机遇论》，第1版问世于1718年，再版于1738年，去世后第2年又发行了第3版(棣莫弗临终委托朋友办理)。每次再版都对内容进行扩充。在第2版中，棣莫弗讨论了“伯努利大数定律”，并在雅各布研究的基础上，以二项分布的逼近导出了正态概率曲线的数学表达式。

2.正态概率曲线的发现过程

棣莫弗同雅各布一样利用某些二项式系数来求概率。他娴熟地利用无穷级数对$(a+b)^n$各项求和近似方法作了详细论述。从中可以看出，他非常擅长利用无穷级数解决概率问题。棣莫弗首先考虑事件成败具有等可能的情况，即$p=q=1/2$，n为较大偶数时，求二项分布中心项$b(n,1/2,n/2)$的近似表示，然后研究中心项与任意项之比，进而发现了正态概率曲线。

(1)二项分布中心项的近似表示。

18 世纪的概率问题需要求二项式中几项的和，当 n 很大时是一个很困难的问题。雅各布曾研究了这一问题，但其精度不理想。在此基础上，棣莫弗开始了这一问题的研究。他假设 $(1+1)^n$ 中的 n 很大，考虑其中心项与所有项的和之比。

1721 年，棣莫弗得到如下结果：

设 $n=2m$（n 为整数），当 $n\to\infty$ 时，

$$b(m)\sim 2.1682(1-1/n)^n/\sqrt{n-1}=2T(1-1/n)^n/\sqrt{n-1},$$

其中 $\ln T=\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots=\frac{B_2}{1\cdot 2}+\frac{B_4}{3\cdot 4}+\frac{B_6}{5\cdot 6}+\frac{B_8}{7\cdot 8}+\cdots$，这里 B_i 是伯努利数。

在推导过程中，棣莫弗先将 $b(m)$ 恒等变形为

$$b(m)=\frac{(2m)!}{m!m!}(\frac{1}{2})^{2m}=\frac{2m(2m-1)\cdots(m+1)2^{-2m}}{m(m-1)\cdots 2\cdot 1}$$
$$=\frac{(m+1)(m+2)\cdots(m+(m-1))(m+m)}{(m-1)(m-2)\cdots(m-(m-1))m}2^{-2m}=2^{-2m+1}\prod_{i=1}^{m-1}\frac{m+i}{m-i},$$

可得

$$\ln b(m)=(1-2m)\ln 2+\sum_{i=1}^{m-1}\ln\frac{1+i/m}{1-i/m}.$$

再由

$$\ln\frac{1+x}{1-x}=2\sum_{k=1}^{+\infty}\frac{x^{2k-1}}{2k-1},|x|<1,$$

可将上式化为

$$\ln b(m)=2\sum_{i=1}^{m-1}\sum_{k=1}^{+\infty}\frac{1}{2k-1}(\frac{i}{m})^{2k-1}=2\sum_{k=1}^{+\infty}\frac{1}{(2k-1)m^{2k-1}}\sum_{i=1}^{m-1}i^{2k-1}.$$

令 $t=\frac{m-1}{m}$，利用伯努利自然数幂求和公式

$$\sum_{i=1}^{n}i^m=\frac{n^{m+1}}{m+1}+\frac{n^m}{2}+\sum_{i=1}^{[m/2]}\frac{1}{2i}B_{2i}C_m^{2i-1}n^{m-(2i-1)},$$

得

$$\ln b(m)=2(m-1)\sum_{k=1}^{+\infty}\frac{t^{2k-1}}{2k(2k-1)}+\sum_{k=1}^{+\infty}\frac{t^{2k-1}}{2k-1}+2\sum_{k=1}^{+\infty}\sum_{i=1}^{k-1}\frac{B_{2i}}{(2k-1)2im^{2i-1}}C_{2k-1}^{2i-1}t^{2k-2i}.$$

这样，右边第二项形成的幂级数其和函数为 $1/[2\ln(2m-1)]$，而第一项对从 0 至 t 积分可得类似于第二项的幂级数，其和函数为 $(2m-1)\ln(2m-1)-2m\ln m$，将第三项改变其求和次序，为

$$L=2\sum_{i=1}^{+\infty}\frac{B_{2i}}{2im^{2i-1}}\sum_{k=i}^{+\infty}\frac{1}{2k-1}C_{2k-1}^{2i-1}t^{2k-2i},$$

而

$$\sum_{k=i}^{+\infty}\frac{1}{2k-1}C_{2k-1}^{2i-1}t^{2k-2i}=\sum_{k=i}^{+\infty}\frac{1}{2i-1}C_{2k-2}^{2i-2}t^{2k-2i}=\frac{1}{2i-1}\sum_{j=1}^{+\infty}C_{2i-2+2j}^{2i-2}t^{2j}$$

$$=\frac{1}{2i-1}\frac{1}{2}[(1-t)^{-(2i-1)}+(1+t)^{-(2i-1)}]=\frac{1}{2(2i-1)}[m^{2i-1}+(1+\frac{m-1}{m})^{-(2i-1)}],$$

故

$$\lim_{m\to+\infty}L=\frac{B_{2i}}{2i(2i-1)},$$

有

$$\ln b(m)\sim(2m-\frac{1}{2})\ln(2m-1)-2m\ln(2m)+\ln 2+\sum_{i=1}^{+\infty}\frac{B_{2i}}{2i(2i-1)}$$

因棣莫弗无法算出右端之和，就只好取其前 4 项近似求和，得

$$\ln b(m)\sim\ln 2+\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}=0.7739,$$

又 $e^{0.7739}=2.1682$，故得出 $b(m)$ 的近似值为

$$b(m)\sim 2.1682(2m-1)^{2m-1/2}(2m)^{-2m}=2.1682(1-1/2m)^{2m}/\sqrt{2m-1}$$

$$=2.1682(1-1/n)^{n}/\sqrt{n-1}$$

(2)斯特林公式的证明。

1725 年，棣莫弗的朋友斯特林(James Stirling，1692~1770)做出了 $b(m)$ 的两个级数表达式为

$$[b(m)]^2=\frac{1}{\pi(2m-1)}(1+\frac{1}{4(m+3/2)}+\frac{9}{32(m+3/2)(m+5/2)}+\cdots),$$

$$[b(m)]^{-2}=\pi m(1+\frac{1}{4(m+1)}+\frac{9}{32(m+1)(m+2)}+\cdots),$$

这是 π 第一次被引到无穷级数中，令 $n\to+\infty(n=2m)$，在上式右边只取主项 1，则有

$$b(m)\sim\sqrt{2/\pi m}$$

棣莫弗得知斯特林的结果后，利用沃利斯(J.Wallis，1616~1703)1655 年所得结论

$$\lim_{n\to+\infty}[\frac{2\cdot4\cdot6\cdots2n}{1\cdot3\cdot5\cdots(2n-1)}\sqrt{\frac{1}{2n+1}}]=\sqrt{\frac{\pi}{2}},$$

并将 $b(m)$ 恒等变形为

$$b(m)=\frac{1\cdot3\cdot5\cdots(2m-1)}{2\cdot4\cdot6\cdots2m},$$

则立即得出

$$b(m)\sim\sqrt{2/\pi m}\qquad(1\text{-}4)$$

1730 年，棣莫弗在其《分析杂论》中给出了对较大的 n，有

$$m!=\sqrt{2\pi}m^{m+1/2}e^{-m}\exp(\frac{1}{12m}-\frac{1}{360m^3}+\cdots)$$

其证明如下：

因为

$$m^m/m! = \prod_{i=1}^{m-1}(1-i/m)^{-1},$$

$$\ln(m^m/m!) = \sum_{i=1}^{m-1}\sum_{k=1}^{+\infty}\frac{1}{k}(\frac{i}{m})^k = \sum_{k=1}^{+\infty}\frac{1}{km^k}\sum_{i=1}^{m-1}i^k,$$

利用伯努利自然数幂求和公式，上式右边则成为

$$(m-1)\sum_{k=1}^{+\infty}\frac{t^k}{k(k+1)}+\frac{1}{2}\sum_{k=1}^{+\infty}\frac{t^k}{k}+\sum_{k=1}^{+\infty}\sum_{r=1}^{[k/2]}\frac{1}{k}C_k^{2r-1}\frac{B_{2r}}{2rm^{2r-1}}t^{k-2r+1},$$

这里$t=(m-1)/m$，类似于上面的积分运算及幂级数求和，则得

$$m-1-\frac{1}{2}\ln m+\sum_{k=1}^{+\infty}\frac{B_{2k}}{2k(2k-1)}(1-m^{1-2k})$$

$$\sim m-1-\frac{1}{2}\ln m+\sum_{k=1}^{+\infty}\frac{B_{2k}}{2k(2k-1)}m = m-1-\frac{1}{2}\ln m+1-\frac{1}{1+1}\ln(2\pi)$$

$$= m-\frac{1}{2}\ln m-\frac{1}{2}\ln(2\pi),$$

有

$$\ln(m^m/m!)\sim m-\frac{1}{2}\ln m-\frac{1}{2}\ln(2\pi)$$

若略去随$m\to+\infty$而趋于 1 的因子，则得今日的形式

$$m!\sim\sqrt{2\pi}m^{m+1/2}e^{-m} \tag{1-5}$$

利用式 1-5 则易得式 1-4。这最后一步工作是由斯特林完成的。棣莫弗曾说："我中断了这一步的研究，后来我值得尊敬的、有学问的朋友斯特林先生做了进一步的探讨和研究，找到了$C=\sqrt{2\pi}$"。斯特林公式不但有很高的理论价值，而且还可以得出一些较精确的数值估计。随着m的增加，式 1-5 两边之差可以任意小，而当m很小时，斯特林逼近很精确。

(3)正态分布和极限定理。

1733 年 11 月 12 日，棣莫弗将其一篇 7 页的论文送给了几位朋友。后来，棣莫弗听取朋友的意见做了修改，又增加一些内容，收录在《机遇论》第 2 版 235~243 页。正是在这篇文章中，他第一次导出了正态概率曲线的表达式。其中研究了二项分布中心项与任意项之比，即

$$b(m)/b(m+d),$$

得出

$$\ln[b(m)/b(m+d)]\sim(m+d-1/2)\ln(m+d-1)+(m-d+1/2)\ln(m-d-1)$$
$$-2m\ln m+\ln\frac{m+d}{m}$$

上式应有限制$d/(2m)\to 0$。这一点棣莫弗虽未明确指出，但在论证时，他没有超出这个限制。

借此，棣莫弗证明了当 $n \to +\infty$ 时，有

$$b(m)/b(m+d) \sim \exp(-2d^2/n)$$

若限制 $|d| \le c\sqrt{n}$ ($c>0$ 为常数)，则由式 1–4 可得

$$b(m+d)=b(m)e^{-\frac{2d^2}{n}}=\frac{2}{\sqrt{2\pi n}}e^{-\frac{2d^2}{n}}$$

据此，棣莫弗指出：如果视二项式展开式的各项为一系列竖直线段的长度，把这些线段摆在同一直线上方且与之垂直，那么线段的上端点将描绘出一条曲线。由此得到的曲线具有两个拐点，它们分别位于最大项对应点的两侧。该曲线就是今日的正态概率曲线，进而棣莫弗又得到以下近似公式

$$\sum_{d=0}^{k}P(X=m+d)\approx\frac{2}{\sqrt{2\pi n}}\int_0^k e^{-2t^2/n}dt\,,$$

即在 n 次独立重复试验中事件出现 m 次的概率之期望值满足

$$\lim_{n\to+\infty}P\left(a<\frac{m-n/2}{\sqrt{n}/2}<b\right)=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\,, \qquad (1\text{–}6)$$

这里 a,b 为任意实数，且 $a<b$。

可以看到，正是上式揭示了二项分布与正态分布间的关系，从而将离散型随机变量与连续型随机变量建立了联系。

1774 年，拉普拉斯首先证明了

$$\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx=\sqrt{2\pi}\,,$$

并对棣莫弗的结果进行推广，建立了中心极限定理较一般的形式，即今天的棣莫弗–拉普拉斯中心极限定理：

若随机变量 $\xi \sim B(n,p)$，则对任何有限实数 a,b ($a \le b$)(或 $a\to-\infty$，$b\to+\infty$)，有

$$\lim_{n\to+\infty}P\left(a<\frac{\xi-np}{\sqrt{np(1-p)}}<b\right)=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

再有，若 $0<p<1$，$q=1-p$，则对任意给定的实数 $a<b$，当 k 满足 $(k-np)/\sqrt{npq}\in(a,b)$ 时，下述极限一致成立

$$\lim_{n\to+\infty}\frac{C_n^k p^k q^{n-k}}{(2\pi npq)^{-1/2}\exp\left(-\frac{(k-np)^2}{2npq}\right)}=1\,,$$

这称之为局部极限定理。当 n 很大，$|k-np|$ 相对于 $\sqrt{npq}$ 不很大时，则可用

$$(2\pi npq)^{-1/2}\exp\left(-\frac{(k-np)^2}{2npq}\right)$$

来近似计算 $C_n^k p^k q^{n-k}$ 之值。虽然，前一表达式看起来复杂，但借助计算器则能迅速求出其值。

3.科学历史评价

利用棣莫弗所得结果可以解决一大类实际应用问题。若某奖券的量很大，其中奖率为 5% ，如果要保证中奖的概率达 90% ，至少应购买多少张奖券？类似地，其方法也可以应用于保险业中。棣莫弗将其结果大量用于诸如此类的问题。这就初步显示了概率论这一数学分支的广泛应用性。

就伯努利与棣莫弗研究的问题而言，伯努利讨论的是当试验次数趋于无穷大时频率的极限行为，而棣莫弗研究的则是有利事件出现次数的相关变量之极限分布，但二者又是相通的。若将式 1–6 变形为

$$P(|\frac{X}{N}-\frac{1}{2}|\le\frac{C}{\sqrt{N}})=\int_{-2C}^{2C}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx.$$

因为这里 C 可以充分大，这就比雅各布更为简洁地证明了伯努利大数定律。在尚无方差概念的当时，其证明的确匠心独具。此外，其精度也有所提高。如对雅各布要求进行 25 550 次试验的情形，而棣莫弗仅需 6 498 次即可。

棣莫弗的结论还否定了过去学术上的两种极端意见，得出观察值的平均值之精度与观察次数 N 的平方根 $\sqrt{N}$ 成正比，这是认识自然的一个重大进展。同时，棣莫弗也看到 $\sqrt{N}$ 的重要地位，特引进“模”这个概念，后来被“标准差”取而代之。

在《机遇论》中，棣莫弗设计了各种新的方法，对各种类似的点数问题、换球问题、年金问题等都进行了系统的研究。他一反通常程序，而从概率原理去推导排列组合公式。他明确提出了统计独立性和条件概率的概念，还为概率论发展了一套普遍的概率符号，并称之为“新代数”。他得到了泊松分布的一种特殊情形，并将母函数用于对正态分布的讨论。其《年金论》不仅改进了以往关于人口统计的方法，而且在假定死亡率所遵循的规律以及银行利息不变的前提下，导出了计算年金的公式，其内容被人奉为经典。由此可见，棣莫弗致力于将概率论应用于人文、社会科学研究中。这在概率论发展中起到了承前启后的作用。

对于棣莫弗自己来说，他认为解决了这样的哲学问题：在人们以为纯粹偶然的事件中，可以寻找其规律和必然。正如他在《机遇论》第 3 版所言，尽管机会具有不规则性，由于机会无限多，随着时间的推移，不规则性与秩序性相比将显得微不足道。他强调，这种秩序自然是从“固有设计中”产生出来的。棣莫弗对他在概率论方面的研究成果感到非常满意，在其《机遇论》的序言中写道：

“在我刚开始解决机会游戏问题时，看不到光明所在，因为蒙特摩先生在他的书中曾给出了这一问题的解决方法，可他仅仅压了三个输或赢的赌注，并且通过假设进一步限制了冒险者之间的一种平等技巧，况且他也没有给出证明的方法，我是努力冲破这一切的。”

时至今日，正态分布牢固地占据了统计分析和概率论的主导地位，已广泛应用于天文及测地数据的分析、社会统计、生物统计、教育统计、体育统计及卫生统计等。追本溯源，棣莫弗的开创之功实不可没。

1.7 蒲丰和贝叶斯的概率论工作

蒲丰(G.L.Buffon，1707~1778)是法国数学家、自然科学家。1707 年 9 月 7 日生于蒙巴

尔；1788 年 4 月 16 日卒于巴黎。

蒲丰 10 岁时在第戎耶稣会学院读书，16 岁主修法学，21 岁到昂热转修数学，并开始研究自然科学，特别是植物学。1733 年当选为法国科学院院士，1739 年任巴黎皇家植物园园长，1753 年进入法兰西学院。1771 年接受法王路易十四的爵封。

蒲丰于 1740 年翻译了牛顿的《流数法》，并探讨了牛顿和莱布尼茨发现微积分的历史。

蒲丰还以研究自然博物史著称，他集多年研究成果编成巨著《自然史》（44 卷，蒲丰生前出版了 36 卷，后 8 卷由他的学生完成）。他是第一个对地质史划分时期的科学家，他还首次提出太阳与彗星碰撞产生行星的理论。

蒲丰是几何概率的开创者，并以蒲丰投针问题闻名于世，发表在其 1777 年的论著《或然性算术试验》中。其中首先提出并解决下列问题：把一个小薄圆片投入被分为若干个小正方形的矩形域中，求使小圆片完全落入某一小正方形内部的概率是多少，接着讨论了投掷正方形薄片和针形物时的概率问题。这些问题都称为蒲丰问题。

其中投针问题的典型模型可述为：

设在平面上有一组平行线，其距都等于a，把一根长$l < a$的针随机投上去，则这根针和一条直线相交的概率是$2l/\pi a$。

这里的随机是指针的中心的落点与针的方向都是等概率的，而且中心落点与针的方向无关。

令ρ表示针的中心与平行线的最近距离 ，θ表示针与平行线的夹角 ，于是相交时(图 1-3)

$$\rho < \frac{l}{2}\sin\theta, 0 < \theta < \pi .$$

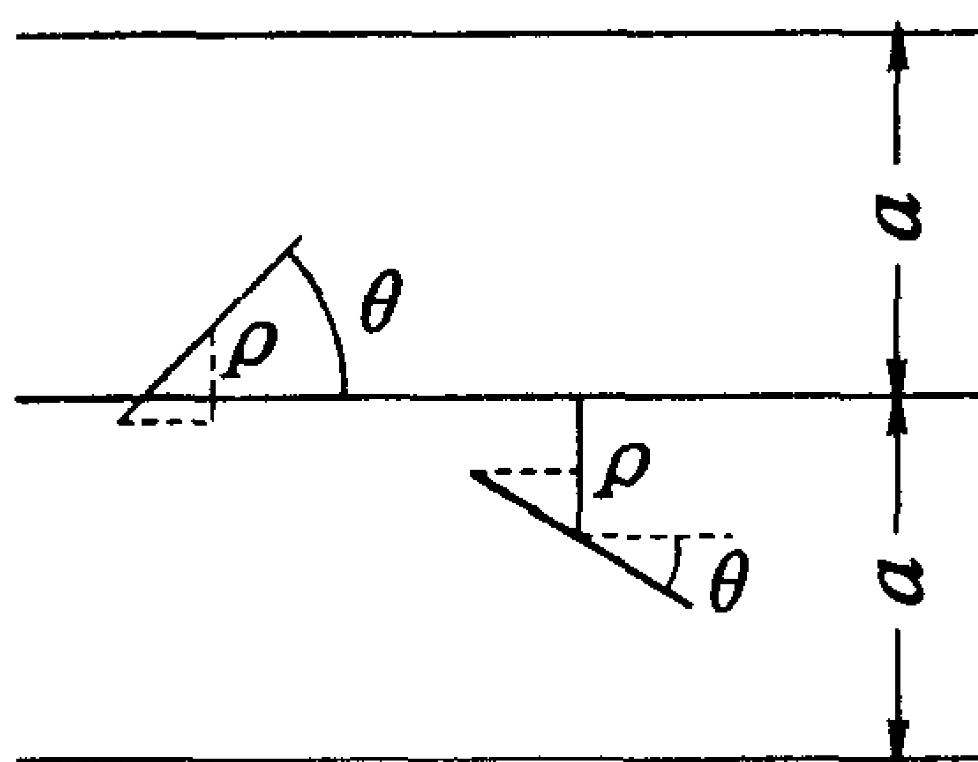


图 1-3

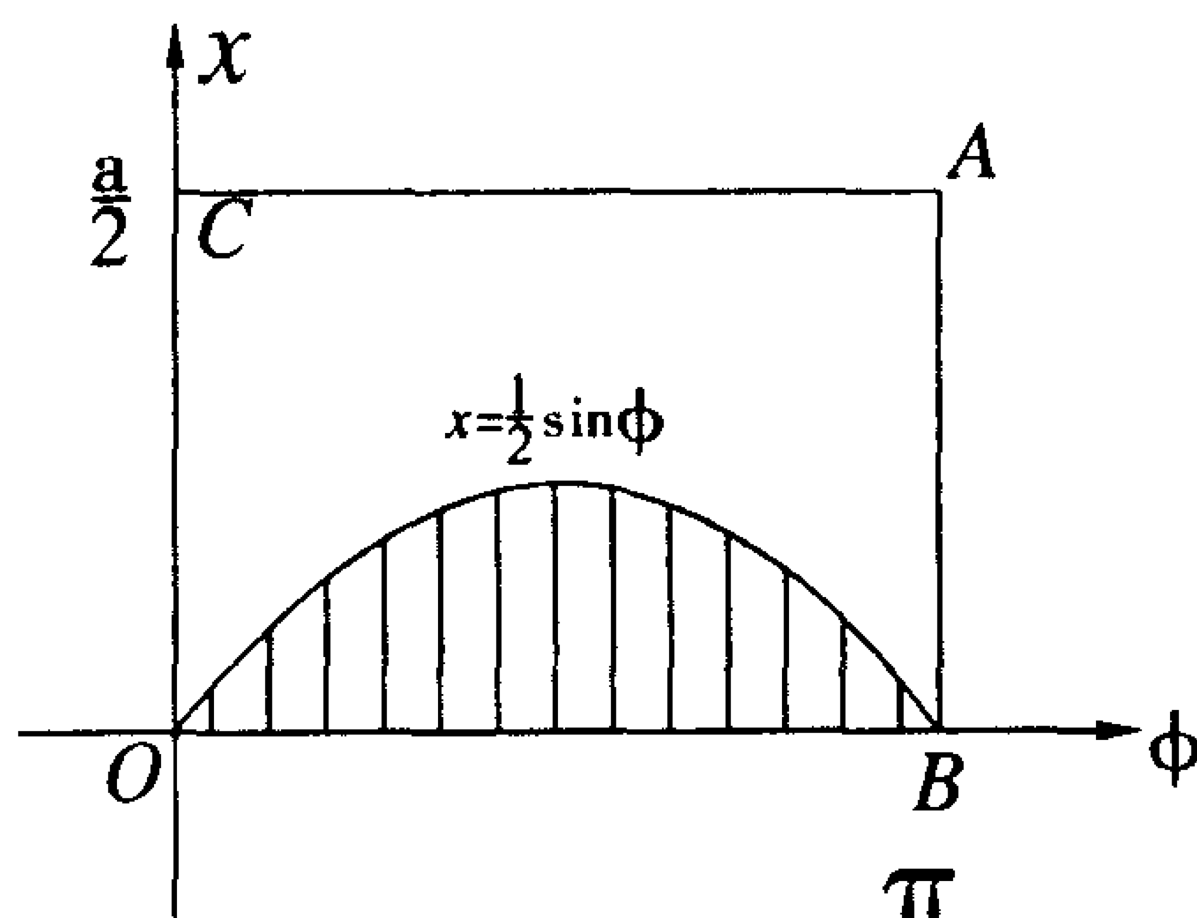


图 1-4

在直角坐标系$o\theta\rho$中，考虑矩形$ABOC$(图 1-4)，此矩形内每一点(ρ,θ)与针落在平面上的方位一一对应。阴影区内的点对应针与平行线相交，于是所求概率为

$$p = \frac{\text{阴影区面积}}{\text{矩形面积}} = \frac{\frac{l}{2}\int_0^{\pi}\sin\theta d\theta}{\frac{a}{2}\pi} = \frac{2l}{\pi a}$$

这是数学史上古典概率论中几何概率的一个精彩实例。

由于$\pi = \dfrac{2l}{pa}$，只要求得p，则可以求出π。由于通过他的投针试验法可以利用很多

次随机投针试验算出π的近似值，所以特别引人瞩目。1850 年，瑞士数学家沃尔夫在苏黎士，用一根长 36mm 的针，平行线间距为 45mm，投掷 5 000 次，得$\pi \approx 3.1596$。1864 年，英国人福克投掷了 1 100 次，求得$\pi \approx 3.1419$。1901 年，意大利人拉泽里尼(Lazzerrini)投掷了 3 408 次，他统计出针与平行线相交的次数m，得到了准确到 6 位小数的π值。

这就告诉我们，可以通过大量重复抛针，由上述公式求出π的近似值。这就为我们开辟了一条完成某些计算任务的途径，这种方法称为蒙特卡罗(Monte Carlo)随机模拟法。由于大量重复试验可以通过计算机来模拟实现，所以 Monte Carlo 法现今已获得了广泛的运用，例如可以用它计算一些原函数不易求出的定积分等。

1812 年，在《分析概率论》中拉普拉斯推广了蒲丰的模型：

两组正交等距平行线，一组距离为a，另一组距离为b，针长$l < \min(a,b)$，则针与任一直线相交的概率为

$$p = \frac{2l(a+b)}{\pi ab},$$

当$b \to +\infty$(或$a \to +\infty$)时，即为蒲丰的结果。

贝叶斯(Bayes，1702~1761)是英国数学家。1702 年生于伦敦，1761 年 4 月 17 日卒于坦布里奇韦尔斯。

贝叶斯是一位自学成才的数学家。曾助理宗教事务，后来长期担任坦布里奇韦尔斯地方教堂的牧师。1742 年，贝叶斯被选为英国皇家学会会员。

在关于微积分基础的论战中，贝叶斯也发表过文章，为了反对贝克莱主教对微积分的攻击，他 1736 年发表了《流数术学说入门》。

1763 年，贝叶斯发表《论机会学说问题的求解》中，提出了一种归纳推理的理论，其中的“贝叶斯定理（或贝叶斯公式）”给出了在已知结果E后，对所有原因C计算其条件概率(后验概率)$P_E(C)$的公式，可以看作最早的一种统计推断程序，以后被一些统计学者发展为一种系统的统计推断方法，称为贝叶斯方法。采用这种方法作为统计推断所得的全部结果，构成贝叶斯统计方法的内容。贝叶斯统计在理论上的进展以及它在应用上的方便和效益，使其观点为许多的人所了解，并对一些统计学者产生吸引力。而认为贝叶斯方法是唯一合理的统计推断方法的统计学者，形成数理统计学中的贝叶斯学派。如今在概率论、数理统计学中以贝叶斯姓氏命名的有贝叶斯公式、贝叶斯风险、贝叶斯决策函数、贝叶斯决策规则、贝叶斯估计量、贝叶斯方法、贝叶斯统计等。

第2章 近代概率论——分析概率论

从1812年至20世纪初，概率论的研究主要采用分析方法，如特征函数、微分方程、差分方程等，这一时期称为分析概率论时期，开创了近代概率论时代。拉普拉斯总结了古典概率论，并使它发展到新的历史阶段，他的著作《分析概率论》就是承上启下的代表作。

2.1 拉普拉斯决定论的成因及其历史地位

拉普拉斯(P.S.Laplace，1749~1827)是法国数学家、天文学家、物理学家。1749年3月23日生于博蒙昂诺日，1827年3月5日卒于巴黎。

拉普拉斯家境贫寒，靠邻居的周济才得到读书的机会。16岁时进入开恩大学，并在学习期间写了一篇关于有限差分的论文。在完成学业之后，他带着介绍信从乡下到巴黎去求见大名鼎鼎的达朗贝尔，荐书投去，杳无音讯，因为达朗贝尔对于只带着大人物的推荐信的年轻人不感兴趣。拉普拉斯并不气馁，随即写了一篇阐述力学一般原理的论文，求教于达朗贝尔。由于这篇论文异常出色，达朗贝尔为其才华所感，欣然回了一封热情洋溢的信，信中写道“拉普拉斯先生，你看，我几乎没有注意你那些推荐信，你不需要什么推荐。你已经更好地介绍了自己。对我来说这就够了，你应该得到支持。”达朗贝尔还很高兴地当了他的教父，并介绍他去巴黎陆军学校任教授。拉普拉斯事业上辉煌时期便从此开始。1773年被选为法国科学院副院士；1783年任军事考试委员，并于1785年主持对一个16岁的唯一考生进行考试，这个考生就是后来成为皇帝的拿破仑（Nopoleon）；1785年当选为法国科学院正式院士；自1795年以后，他先后任巴黎综合工科学校和高等师范学校教授；1816年被选为法兰西学院院士，一年后任该院主席。他还被拿破仑任命为内政部长，元老议员并加封伯爵。拿破仑下台后，路易十八（Louis XⅧ）重登王位，拉普拉斯又被晋升为侯爵。

拉普拉斯才华横溢，著作如林，在青年时代就发表了一系列的论著。24岁当选为法国科学院副院士，科学院在一份报告中曾这样评价他：还没有任何一位像拉普拉斯这样年轻的科学家能在如此众多如此困难的课题上，写出如此大量的论文。

拉普拉斯的研究领域是多方面的，有天体力学、概率论、微分方程、复变函数、势函数理论、代数、测地学、毛细现象理论等，并有卓越的创见。他是一位分析学的大师，把分析学应用到力学，特别是天体力学，获得了划时代的结果。他的代表作有：《宇宙体系论》、《分析概率论》、《天体力学》，素有“法国的牛顿”之称。

拉普拉斯在政治上是一个机会主义者。在法国大革命时期，随着政局的动荡、改朝换代，他也随波逐流，反复不断地扮演了共和派与保皇派的双重角色，他机灵到能够使敌对的双方在不论哪一方上台掌权时，都相信他是自己的一个忠诚的支持者，因此每次改宗后他都能获得更好的差使和更大的头衔。为此有人把他比作英国文学作品中的假圣人布雷牧师。拿破仑在流放期间说过：“拉普拉斯是第一流的数学家，但事实很快表明他不过是一个平庸的行政官员，……他把无穷小精神带进了政府之中。”拉普拉斯的另一个缺点是：在他的著作中，他常常完全不提前人和同时代人的论述与功绩，给人的印象是其著作中的

思想似乎完全出自于他本人。例如，他在《天体力学》中不声不响地从拉格朗日那里取用了位势概念，并把这一概念用得十分广泛，以致从他那时起，势论中的基本微分方程被人称作拉普拉斯方程。他在《分析概率论》中，引用别人的成果也不提及别人的名字，而是把它们同自己的成果混在一起。他的这些品格遭到了后人的非议。

拉普拉斯虽有上述缺点，但作为一个科学家，在席卷法国的政治变动中，包括拿破仑的兴起和衰落，都并未显著地影响他对科学的研究。另外他也能慷慨帮助和鼓励年轻的一代。例如，化学家盖・吕萨克（G.Lussac）、旅行家和自然研究者洪堡尔晓（Humboldt）、数学家泊松（Poisson）、柯西都曾得到过他的帮助和鼓励。

他学识渊博，但学而不厌。他留下了牛顿式的遗言："我们知道的是微小的，我们不知道的是无限的。" 他曾说："自然的一切结果都只是数目不多的一些不变规律的数学结论。"他还强调指出："认识一位巨人的研究方法，对于科学的进步……，并不比发现本身更少用处。科学研究的方法经常是极富兴趣的部分"。

在数学中以他的姓氏命名的有：拉普拉斯变换、拉普拉斯定理、拉普拉斯方程、拉普拉斯函数、拉普拉斯积分、拉普拉斯极限公式、拉普拉斯算子、拉普拉斯展开、拉普拉斯向量、拉普拉斯序列、拉普拉斯分布、拉普拉斯—傅里叶核等，而其中以他的姓氏命名的变换、定理、方程等有多种。

决定论与非决定论是处理事物之间因果关系的两种不同理论。决定论认为各种事物相互之间具有规律性联系；非决定论否认这种规律性的联系，是与决定论相对立的因果观。这两种理论对自然界的描述，究竟谁为根本?两个多世纪以来科学家们和哲学家们一直争论不休。在这场争论中，拉普拉斯被推举为机械决定论的代表，他的这段名言被反复引用：

"我们应当把宇宙目前的状态看做是它先前状态的结果，并且是它以后状态的原因。暂时设想有一位神灵，它能够知道施于自然界的所有作用力以及自然界所有组成物的各自位置，它并且能够十分广泛地分析这些数据，那么，它就可以把宇宙中最重物体和最轻原子的运动，均纳人同一公式。对于它，再没有什么事物是不确定的，未来和过去一样，均呈现在它的面前"。

人们以此作为判定拉普拉斯是机械决定论者的证据，并将这种决定论称为拉普拉斯决定论。

拉普拉斯生活于 1749~1827 年，他既是当时著名的数学家和天文学家，又是政治上的风云人物。他一方面发展和完善了牛顿力学，另一方面又对那个时代的概率成果进行了系统的总结，成为当时概率论的集大成者。概率论是一门研究或然现象的学科，机械决定论则否认或然现象的存在，二者在拉普拉斯身上的水乳交融是否使他看到了规律内在的某种不确定性和偶然性因素的存在呢?在机械决定论思想的支配下，他又为何能提出包含运动变化观点的"星云假说"呢? 我们力图在对拉普拉斯决定论的成因进行考察的同时，结合拉普拉斯本人的著作，尝试从一个侧面给出上述问题的答案，并对拉普拉斯决定论的历史地位作出客观的评述。我们认为，这对于探讨学术界的决定论与非决定论之争，对于探索自然科学与哲学之间的关系，对于科学史的研究，都具有一定的理论和实际意义。

1.拉普拉斯生活的社会历史背景对其哲学思想的影响

拉普拉斯生活的时代是资产阶级革命以法国为中心的时代。为从思想上、理论上彻底清除基督教神学，18 世纪的法国爆发了一场以宣传牛顿力学体系和机械决定论为中心内容的法国启蒙运动。这场运动直接影响了拉普拉斯的哲学思想。

我们知道，牛顿力学的中心内容是牛顿三定律和万有引力定律。牛顿三定律是机械运动认识的系统总结，在实践中得到了广泛应用。而万有引力定律则讨论了运动的原因，定量地解释了地球与木星由于它们的自转而中部突出的问题、解释了由于地球与月球之间的引力而发生的潮汐现象，以及彗星在太阳系中的运行轨道，并且预见了后来发现的天王星、海王星等。对于自然现象这种惊人的明察，使牛顿力学成为科学界的最高权威和各门物理学的理论基础。

力学所取得的巨大成就，使得牛顿企图把一切自然现象都归结为力学现象，把一切运动都归结为机械运动。他把一切自然现象用“力”来联结，并且用力学原理来论证整个自然过程。他认为：

“自然哲学的目的在于发现自然界的结构和作用，并且尽可能把它归结为一些普遍的法则和一般定律——用观察和实验来建立这些法则，从而导出事物的原因和结果”。

“哲学的全部任务看来就在于从各种运动现象来研究各种自然之力，尔后用这些力去论证其他现象”。

受牛顿力学及其哲学思想的影响，人们片面地把力学规律看做是唯一的、最普遍的规律，并企图以此去解释整个自然界，于是产生了机械自然观。

机械自然观在因果关系上的具体表现就是机械决定论。由于牛顿力学得出了严格的用数值来表示的机械因果性公式，人们根据它可以精确预言机械运动的结果，于是这种机械的单义决定论就被看成了决定论的唯一形式，这实际上就是认为整个自然界是由机械决定的因果必然链条构成的。正如牛顿所说：“自然界的其他一切现象，全可以根据力学原理，用相似原理，一一演绎出来”。

正是由于牛顿力学体系的这些特点，使其顺应了资产阶级革命欲摆脱宗教神学束缚的要求。到了 18 世纪中叶，启蒙运动进入高潮，群星灿烂，人才辈出，大批思想界巨人纷纷涌现。他们不辞劳苦，传播牛顿力学和机械唯物论思想，开启人们的头脑，就在这样一种年代里，拉普拉斯呱呱坠地了。

启蒙运动的重要成果之一是 18 世纪 50~70 年代出版的法国《百科全书》，编撰者来自各个方面，其中著名的人士之一达朗贝尔，一直协助狄德罗担任数学编辑。在思想上，达朗贝尔与其他启蒙者一样，主张理性和自然权利，极力推崇牛顿力学。26 岁时，他发表了动力学的一部基础性著作《论动力学》(1743)，书中有著名的“达朗贝尔原理”，这实际上是牛顿第二定律的另一种表达形式，它把动力学问题简化成了静力学问题。他还认为，牛顿的第三运动定律对于运动的物体和固定不动的物体都适用。紧接着，他发展了微分方程，用数学分析的手段进一步丰富和完善了牛顿力学体系。总之，“达朗贝尔的著作继承了牛顿的发现”，他在思想上和学术上都带有牛顿力学的烙印。

拉普拉斯这位在法国启蒙运动中降生的农家子弟，19 岁时离家去巴黎谋求职业。举目无亲的拉普拉斯受人之荐，拜访了达朗贝尔，想得到达朗贝尔的提携，却遭到了达朗贝尔的拒绝。聪明的拉普拉斯把他的研究成果写成论文再次递交给达朗贝尔，正直的达朗贝尔被拉普拉斯的才能打动了。他推荐拉普拉斯做了巴黎军事学校的数学教授。从此，拉普拉斯与达朗贝尔的关系日益密切，后来达朗贝尔还做了拉普拉斯的教父。达朗贝尔的思想深深影响着拉普拉斯，达朗贝尔用他的“著作培养了后继者拉普拉斯”。在达朗贝尔的机械自然观熏陶下，拉普拉斯目睹了牛顿力学的巨大成就，于是他决心把牛顿力学引入天文学。

因此，我们可以说，拉普拉斯是在机械自然观的指导下，踏上了科学研究的征途，在探寻事物因果关系上，他采用了机械决定论的观点也就不足为奇了。

2.拉普拉斯的知识背景对其哲学思想的影响

在把牛顿力学应用于天文学方面，拉普拉斯取得了巨大的成功，使理论天文学产生了飞跃的发展：1773 年他着手把牛顿力学应用于整个太阳系，他探讨了土星轨道为什么连续膨胀，而木星轨道则不断收缩这个特别困难的问题；他还发现了行星平均运动的不变性。拉普拉斯致力于研究行星之间的摄动，1780 年证明行星轨道的偏心率和倾角总保持很小和恒定，能自动调整，即摄动效应是守恒和周期性的，既不会累积也不会消解。1787 年，他发现月球的加速度同地球轨道的偏心率有关，从理论上解决了太阳系动态中观测到的最后一个反常问题。1796 年，他的著作《宇宙体系论》问世，书中提出了对后世有重大影响的关于太阳系起源的星云假说。1789~1827 年间，他发表了《天体力学》五卷集，总结了他把数学和力学定律应用于天文学所取得的成果。

拉普拉斯在天文学上的巨大成功，使他获得了“法国的牛顿”的殊荣。纵观拉普拉斯的天文学成就，无不与牛顿的力学思想息息相关。牛顿力学是拉普拉斯在自然科学上成功的基础，同时也是他的机械决定论思想形成的关键之所在。其中，万有引力定律对其哲学思想的影响最为巨大。例如在《宇宙体系论》一书中，拉普拉斯在谈到行星轨道、行星和彗星运行所遵守的定律及周期差数、月球和木星卫星的周期差数、二分点的岁差，地轴转动、月球极轴的运动和潮汐的涨落等现象时，写道：

“如果考虑到现今没有一个现象不是可以从引力定律得到解说的话，而且考虑到这个定律以很高的精度决定天体在每一瞬间和整个进程里的位置与运动，我们便不怕这个定律为某个还没有观测到的现象所否定，最后，天王星和它的卫星以及新发现的四颗小行星都顺从，而且验证了引力定律；我们不能否认这一切证据，使我们不得不肯定，除了地球的运动和万有引力之外，自然哲学里没有什么更完美的论证了”。

这段话不仅赞美了万有引力定律的成就，更重要的是它高度概括了牛顿力学中的因果律。

在拉普拉斯生活的那个时代，自然科学还处在宏观低速范围内，牛顿力学完全可以解决当时自然科学所提出的一切问题。因此，拉普拉斯把牛顿力学中的那种单链因果关系看成了唯一正确的因果律，并由此推广到自然界和人类生活的各个领域，从而坚定不移地相信：“一切事件，即使是那些由于意义不大而似乎是不符合伟大的自然规律的事件，也都正如太阳的运转那样是自然界的必然结果”。偶然性实际上是不存在的。我们所说的偶然性，是由于我们“对这类事件与整个宇宙系统之间的联系的无知所造成的，它们必将随着知识范围的拓宽而渐渐减少，并且在正确的哲学面前彻底消失”。这与霍布斯、霍尔巴赫等机械唯物主义者的观点是十分相近的。恩格斯在《自然辩证法》中对这类机械决定论者进行了生动的刻画，揭示了这类机械论者看问题的实质所在：把一切都看成必然的，实际上是把必然性降低成了偶然性。

当然，受牛顿力学中单链因果关系影响的拉普拉斯是意识不到这一点的，他信手拈来了一些佐证材料来证实他的观点。他以哈雷彗星的发现为例指出：

“在并不遥远的过去，不寻常的暴雨、严酷的干旱、拖着长长尾巴的彗星、日食和月食、北极光等，由于它们的出现和消失时间间隔很久，人们对此很不理解，因而常常将此看成是天神愤怒的种种表现，似乎这是与自然规律相违背的偶然现象。所以，1456 年彗星

长长的尾巴引起的恐怖遍及欧洲，而土耳其人推翻东罗马帝国(Lower Empire)的迅速成功本来就已使欧洲陷入一片惊慌之中。然而，熟知天文学知识的哈雷(Halley)，通过分析发现，这颗彗星的出现不是偶然的，而是具有铁的必然性，1456 年、1531 年、1607 年和 1682 年出现的彗星是同一颗彗星，并且指出这颗彗星下一次出现的时间应在 1758 年年底或 1759 年年初。知识界以不耐烦的态度等待着它的出现。果然，哈雷的预言得到了证实。”

哈雷彗星的发现显然在拉普拉斯的思想上留下了强有力的印象。紧接着，克莱罗(Clairaut)在分析木星和土星的作用而引起该彗星的摄动时，确定了这颗彗星下一次于 1759 年 4 月初在近日点经过的路径，并且这些也由观察结果确凿地证实了。所有这些，使拉普拉斯相信，人类的心志在天文学上所表现出来的完美，为他的机械决定论思想提供了有力的证据。由此，拉普拉斯用生动的语言得出了他的著名结论，即拉普拉斯的那段名言。

后来，英国天文学家金斯(又译秦斯，James 或 Jeans)曾形象地描述了拉氏的这种观点：世界的过去、现在和未来是早已安排好了的，它就像一个卷着的巨大毛毯，图案的花纹早已织好，只是在随着时间的推移逐渐展开而已。

上述分析表明，拉普拉斯在把牛顿力学应用到天文学的过程中，充分体验到了牛顿力学的奇妙功效，因而在探讨事物的因果关系时，形成了典型的拉普拉斯机械决定论。可见，拉普拉斯的天体力学成就对其哲学思想的影响是巨大的、根深蒂固的。这也充分体现了哲学与自然科学之间的密切关系。在拉普拉斯的科学生涯中，他除了在天体力学方面作出的巨大贡献外，还是概率论的集大成者。拉普拉斯 1774 年发表的一篇论文(“Memoire Surles Suites recurro—recurrentes et Sur leurs usages dans la theorie des hazards”)中对概率下了一个定义：

“一事件的概率等于每种有利情形与其概率的乘积之和除以各种可能情形与其概率的乘积之和，如果各种情形是等可能的，那么事件的概率就等于有利情形的数目除以所有可能情形的数目”。

这一定义在概率的发展史上占有很重要的地位，它是首次为概率所下的定义，是古典概率发展的基础。在数学的或然性理论中，概率代表一个量，拉普拉斯的定义给出了该量的数值。

后来，拉普拉斯探讨了原因的概率。在上述概率定义的基础上，他在 1774 年发表的一篇论文(“Memoire Sur la probabilite des causes par les evenements”)中指出：

“如果一事件是由 n 种不同版，那么由该事件所预计的原因存在的概率就等于由这些原因所预计的概率；每一原因的概率就等于由这些原因所预计的事件的概率除以由所有原因所预计的事件的概率之和”。

除此之外，拉普拉斯还详细探讨了概率在自然科学和道德科学上的应用，并在这些方面发表了大量论文。但到了 1781 年前后，他却突然中止了在概率论方面的研究，转而从事他的巨著《天体力学》的写作工作。他之所以中止概率论的研究，主要是因为在 18 世纪，与牛顿力学相对应的机械自然观达到了鼎盛时期，在这种氛围下，他决心把牛顿力学引入天文学。也就是说，拉普拉斯毕生始终如一地致力于用牛顿力学来研究天体的运动。他在数学方面所取得的成就，一方面是由于他的爱好和天赋，另一个重要的方面，是因为他在进行天文学研究时，需要数学手段的帮助。不论他在数学上做什么，都是为了对他的天文学研究提供帮助。

正如泊松(Poisson)对他的评价那样：

“拉普拉斯则主要把数学看作一个工具，当每一个特殊问题出现时，他就巧妙地修改这个工具，使它适合于该问题……，他是一位伟大的哲学家，试图通过高等数学为自然服务来了解自然”。

傅里叶也认为：

“拉普拉斯把他的全部著作用在一个固定方面，他从来没有偏离过它。冷静沉着地坚持他的观点，一直是他天才的主要特点。在概率论方面的划时代的工作，尽管乍看上去偏离了他的主要兴趣，实际上也是受了他在数学天文学中的需要的启迪。一旦深入了这个理论，他就看出，它对于一切精密的科学都是必不可少的，感到尽他的力量去发展它是有道理的。”

所以，拉普拉斯在他的著作《天体力学》中，在数学上所作的解释是极其粗糙的，他只注重结果。为了避免把复杂的数学论证压缩成简单明潦的形式所带来的麻烦，他常常删去一切论证，只留下结论和一句乐观的评语：“这是显而易见的”。而阅读他的著作的读者则往往要花上几小时、甚至几天的时间，才能看出它是如何“显而易见”的。

所以，尽管拉普拉斯在概率论方面取得了巨大的成就，但他始终没有把它作为他的主要研究对象，故他对概率论本身所进行的哲学探讨并不多。到了 1812 年，他才将他在概率论方面的早期研究成果加以系统化，形成了《分析概率论》一书。该书的出版，使他重新又意识到了对概率进行哲学探讨的必要性。在《分析概率论》与广大读者见面之前，拉普拉斯针对该书的出版指出：我对概率问题很关心。概率论的研究应引起数学家们的重视，因为概率论所获得的一些近似公式有着广泛的应用；同时，概率论问题也应引起哲学家们的注意，因为在一些我们看似偶然的事件中，最终的分析结果却可以得出某种规律性，诸如年金、银行利率、保险金的设置等，都是依赖于大量重复出现的事件的平均结果具有规律性。需要有单独的著作来对概率进行哲学探讨。

也许正是基于这种考虑，拉普拉斯把他早年在师范学校所作的一个演讲进行了重新整理，以《概率的哲学探讨》为题发表了，也许它可以算作是拉氏单独对概率进行哲学探讨的著作。在该书中，他认为：“一个简单的空气分子或者蒸汽分子所描绘的曲线以一种像行星轨道一样确定的方式被受到控制。它们之间仅有的不同只是由于我们的无知所引起的”。

于是，他得出结论：“概率部分地与这种无知有关，部分地与我们的知识有关”。

因此，我们可以看出在拉普拉斯观点的深处包含着这样两个方面的内容：

(1)牛顿力学和宇宙天体运动变化的规律性证明了自然界的发展是有规律可循的，客观事物之间存在着必然的联系。

(2)在自然界的发展变化中，存在着诸多干扰因素。这些干扰因素的存在，使我们无法认清事物的本来面目，因而想当然地认为它们是偶然的。但这些事情的产生也是有原因的，只是恒常规律是由一般性的原因造成的，特殊现象是由特殊原因造成的。恒常规律和特殊现象都反映了因果必然性。也正是由于原因存在着一般原因和特殊原因之别，所以才导致了概率论的产生。

他认为：“在进一步研究之前，掌握‘或然（Chance）’和‘概率(probability)’的意义是很重要的。我们把一件事情看成是或然现象的结果，这是因为我们没有发现事情的内

在规律性。也就是说，我们忽视了产生它们的原因。因此，或然自身并不存在，它只是一个代表我们无知的字眼，除此之外，它不代表任何东西”

所以，尽管拉普拉斯在涉猎概率论这样一门关于或然现象的学科时，似乎也意识到了一些东西，为此他不止在一个地方强调指出概率论的研究应引起哲学家的注意。但由于时代的局限性，有限的思辨能力使他很难借助于概率论提出偶然性与必然性之间的辩证关系。并且，概率论本身也并不给机械因果关系以致命打击，因为概率论研究的虽然是或然现象，但它也表明对于某一类单独看来是偶然的事件，如果把它们联系起来考虑，便可以发现它们的必然性。因此，拉普拉斯并没有因为涉猎概率论而放弃他的机械决定论信仰。

3.拉普拉斯机械决定论的进步意义及其局限性

机械决定论的历史作用是巨大的。机械决定论使科学挣脱了宗教神学的束缚，在科学中清除了封建迷信和非科学的概念。利用机械因果律，探寻客观事物之间的必然联系，在科学上取得了巨大的成功。19 世纪许多自然科学家，如门捷列夫、基尔霍夫、赫尔姆霍兹、开尔文、韦伯、斯托克斯、玻尔兹曼、麦克斯韦、洛仑兹……，都是在机械决定论思想的支配下，作出了重大的科学发现和理论概括。这一历史事实充分说明了机械决定论在科学发展史上的进步作用。

拉普拉斯决定论是自然观层次上的机械决定论，在这种决定论思想的支配下，拉普拉斯在天文学和概率论两方面也都作出了巨大成就。尤其值得一提的是，他于 1796 年发表的《宇宙体系论》一书的“附录七”中，提出的关于太阳系起源的星云假说还包含了丰富的辩证法思想。

他认为太空最初弥漫着巨大的球状星云，炽热并在缓缓自转。根据角动量守恒原理，后来由于冷却而不断收缩时，转动速度便增大，离心惯性力越来越大。在离心惯性力与中心部分的吸引力的作用下，星云逐渐变成扁平的盘状。当离心惯性力与引力相等时，边缘的物质就不再继续收缩而停留在原处，形成一个围绕中心旋转的环，这样的过程可以重复多次，形成若干个气体环。环内密度较大的物质吸引了周围的物质，逐渐形成围绕中心旋转的团块。中心体收缩为太阳，周围的团块冷却后形成行星。

他为什么能提出包含运动变化思想的星云假说呢?究其原因还是因为他受了机械决定论观点的影响，始终不渝地认为任何事物的出现都不是偶然的，都是由一定的原因造成的，同样太阳系的产生也具有铁的必然性。他研究了太阳系三十个天体的运动，从而得出一个结论：太阳系具有明显的规则性。例如，所有的行星都从西向东运动，并基本在一个平面上；卫星运动的方向同行星相同：太阳行星、卫星自转的方向同公转方向相同，并基本在公转轨道的平面上。他的机械决定论信仰以及他在天体力学和概率论方面的成功，使他坚信这种规则性并非出自偶然。为了说明这种规则性的必然根据，他提出了太阳系起源的星云假说。

就拉普拉斯的星云假说自身而言，也是把发展过程只看作量的逐渐变化过程，变化的机制吸引和排斥这种机械力的作用，这并没有超出机械自然观的范围。

然而，在机械决定论思想支配下所提出的星云假说，客观上起了一种什么作用呢?我们知道，恩格斯曾高度评价过拉普拉斯的星云假说，他在《反杜林论》中以兴奋的心情写道：“拉普拉斯以一种至今还没有人超过的方式详细证明了太阳系如何从一个气团发展而来，以后的科学将越来越证实他的观点”。恩格斯认为拉普拉斯的星云假说在僵化的形而

上学自然观上打开了第一个缺口，在辩证法史上占据重要的历史地位。

科学史上，在机械决定论思想支配下提出包含辩证法意义的科学理论，绝非仅此一例。例如，门捷列夫元素周期律，充分体现了从量变到质变的辩证关系，堪称量变质变规律的科学模型，是“一个勋业，这个勋业可以和勒维烈计算尚未知道的行星海王星轨道的勋业居同等地位”。然而，完成这一勋业的门捷列夫却不具备辩证的自然观，他在“把自己的力量贡献给物质的研究时，曾经看出这种研究含有一两个这样的特征或性质：质量，它占有空间并且表现在引力方面，而最清楚和最现实的是表现在重量上；特征，它表现在化学变化中，而最清楚的是表现在化学元素概念中，……在质量和化学元素中间必然有联系。……物质的质量是物质的这样一种性质，物质的其他性质都应取决于它”。正是从质量决定物质的其他一切性质的机械决定论思想出发，才导致了元素周期律的发现。

这充分说明：机械决定论，或自然观层次上的拉普拉斯决定论，在当时的社会历史条件下，并没有中止辩证法及其问题，更没有也不可能成为辩证法的对立物。它是辩证唯物主义自然观的必要准备和前提，作为思维的低级阶段是有其存在的合理性的。就认识过程而言，如果脱离了客观事物各个部分之间的联系，那么，所谓由感性具体→思维的抽象→思维的具体的飞跃也只能是思辨的。就当时的社会历史条件而言，如果不追溯客观事物之间的必然联系，把一切都看成是偶然的，是上帝创造的，就无法清除宗教神学对自然科学的束缚，也就根本谈不上向辩证自然观的过渡。就拉普拉斯及其他一些在机械决定论思想支配下取得了包含辩证思维的科学成就的科学家自身而言，他们思想的最深处往往不自觉地应用了辩证法思想，这是他们取得重大成果的更深刻的内在根据，虽然他们自己并没有意识到这一点。他们尽管不了解偶然性与必然性之间的辩证关系，却并不妨碍他们在科学研究中运用辩证思维。机械决定论思想促使他们探索客观事物之间的因果联系，是他们取得重大科学成果的直接动力，而隐藏在他们的机械决定论思想背后的辩证思想，则是潜在的动力。机械决定论在一定条件下，一定程度上包含着辩证法的观点。

总之，机械决定论的历史作用表现在以下三个方面：

(1)强调事物因果联系的必然性，使自然科学挣脱了神学的锁链。

(2)经典自然科学的形成和发展是与一大批 17~19 世纪的自然科学家的名字连在一起的，而这些科学家都无一例外地受到了机械决定论思想的影响。

(3)就辩证自然观的形成过程而言，机械自然观是辩证自然观的一个不可或缺的准备阶段。

这正是机械决定论或拉普拉斯决定论的进步意义所在。

拉普拉斯决定论最终要借助神灵的帮助。

在拉普拉斯的生涯中，有一段很著名的逸事。拉普拉斯的巨著《天体力学》问世后，他曾将此书呈现给了当时的法国皇帝拿破仑。丹皮尔在他的《科学史》中描述了此次献书的情形：

“有人告诉拿破仑，那本书没有提到上帝的名字。拿破仑是喜欢拿话来难人的，他收到那本书时说：‘拉普拉斯先生，有人告诉我，你写了这部讨论宇宙体系的大著作，但从不提到它的创造者’。拉普拉斯虽是最圆滑的政客，但在他的哲学的每一点上，却有殉道者的坚强不屈的气概，于是他挺直了身子，直率地答道：‘我用不着那样的假设’。拿破仑觉得那个回答很有趣，把这个回答告诉了拉格朗日。拉格朗日说道：‘那是一个美丽的

假设，它可以解释很多东西。’”

恩格斯也曾注意到拉普拉斯的这种回答，他把它看成唯物主义的自然科学家反对神学的例证。他说：“上帝在信仰他的自然科学家那里所得到的待遇，比在任何地方所得到的都更坏。唯物主义只管说明事物，是不理会这些词儿的。只有当那些咄咄逼人的善男信女们要把上帝强加于他们的时候，他们才去理睬它，并且简单地给予回答——也许像拉普拉斯那样说：‘陛下，我不……’”

拉普拉斯的确从天文学中剔除了上帝的存在，他对牛顿导出“上帝第一推动”深为惋惜，他认为牛顿的“上帝第一推动”是牛顿掩饰无知的一种手段。他说：

“对于某些隐晦难知的关系，更聪明的办法是承认我们的无知而不要在追索我们感兴趣的事物的本源时，仅仅为了镇定我们的不安的需要，去用想象的原因说明它们。

这里，我不能自禁地指出牛顿所陷入的歧途，他就在这一点上偏离了他已经应用得很有成效的方法。自从这位大数学家刊布他对宇宙体系与对光的发现以来，他从事另外一种玄想，企图寻找出大自然的创造者赋予太阳系如上述结构的根本动机何在。牛顿在《自然哲学的数学原理》的附录内，对于行星与卫星差不多在同一平面内、接近正圆的轨道上循着相同的方向运行的奇特现象解释后，他还说：‘这些异常奇特的现象完全不是由于机械的原因，因为彗星在天空任何方向上、偏心率很大的轨道上运动，……太阳、行星与卫星的这种美妙的安排，只能是一位全智全能的上帝的创作’。……试追溯人类智慧的发展与其犯错误的历史，我们便会知道‘最后因’常在我们的认识界限前面退却。这些‘最后因’，在牛顿时代，为了解释流星，是安放在大气里的，但被牛顿搬到了太阳系的边界，因此它们在哲学家眼里是愚昧无知的说法”。

拉普拉斯对牛顿的批评是中肯的。然而，饶有兴趣的是，这位在科学上有着“殉道者的坚强不屈的气概”的大科学家，在指出了牛顿所犯的错误的同时，他在探讨事物的因果关系方面，为了说明宇宙的目前状态是它先前状态的结果，又是它以后状态的原因，他也请来了“神灵”。他要设想一位“神灵”，借助于这位“神灵”，可以把宇宙中一切物体的运动“均纳入同一公式”，这样一切就都是确定的了，“未来同过去一样，均呈现在它的面前”，而“人的心智是与此相距无穷之遥远”的。

作为一个唯物论者的拉普拉斯，在探讨事物的因果关系时，为何要请出他很厌倦的“神灵”呢?也许是出于说明问题的方便，但这也可充分看出机械决定论在处理事物因果关系时所存在的缺憾。我们不妨考察一下牛顿引出“上帝”的缘由。众所周知，牛顿的科学思想是以原子论为基础的，而原子论在历史上则是以无神论著称的，所以，我们可以说在牛顿早年的科学生涯中是没有给上帝留下位置的。但是，很遗憾，在牛顿的晚年中，正如拉普拉斯所评价他的那样，牛顿陷入了“歧途”，他晚年的著作中反复提到上帝。当然，这一方面是由于在他生活的时代，保守的宗教势力还很强大，牛顿不得不作出妥协。但还有一个更重要的方面，就是牛顿在机械自然观的指导下，只承认线性因果链，沿着这条线性因果链在追溯机械运动的第一个原因时，牛顿无以作答，只得引出“上帝第一推动”。实际上，牛顿也认为：“上帝是一个代名词，与他的仆人有关，……一个人要证明有一个完美的神，却未同时证明他就是造物主或万物的创造者，则就尚未证明上帝存在，一个永恒的、无限的、全智的和最完美的却无支配权的神，不是上帝，而是自然……”。可见，牛顿在科学上所说的“上帝”，实质上是指他尚未了解的自然，这与拉普拉斯在寻求终极原

因时，不得不假定一位神灵，以便使宇宙中的一切运动“均纳入同一公式”，实际上是一回事。拉普拉斯比牛顿高明的地方仅在于，牛顿把神搬到了太阳系的边缘，而拉普拉斯提出了一个关于太阳系起源的星云假说，从而把神从整个天文学中请了出去。而这种高明并不是因为他放弃了机械决定论信仰，而是因为他比牛顿占据更多的天文学知识。但是，拉氏最终还是没有完全放逐神灵，他把从天文学中请出去的“神灵”又安放在了哲学的探讨之中。因此，这足以说明，单链因果联系的机械决定论贯彻到底，是离不开神灵的帮助的。随着现代自然科学的发展，它必然地让位于新的因果观。

2.2 拉普拉斯的《分析概率论》研究

雅各布·伯努利之后，棣莫弗、拉普拉斯、高斯、泊松等相继对概率论作出了进一步的工作。其中拉普拉斯 1812 年出版的《分析概率论》实现了概率论研究由组合技巧向分析方法的过渡，开创了概率论发展的新阶段。《分析概率论》是对前人及他自己研究成果的全面总结。该书运用 17、18 世纪发展起来的强有力的分析工具处理概率论的基本内容，使以往零散结果系统化。正是在这部著作中，拉普拉斯给出了著名的概率古典定义，导出了原始形式的中心极限定理（后称棣莫弗——拉普拉斯中心极限定理），并将概率论广泛应用于观测误差估计、人口统计、保险等科学与社会问题。《分析概率论》第 2 版（1814）中增加了一个长达 150 页的绪论，同年印成单行本，题为《概率的哲学导论》，论述概率论定义、发展历史、概率计算的一般原理与应用，并阐明了概率论的重要概念——数学期望及其分析计算方法。这两文作为姐妹篇多次一起或单独再版（我们所指的《分析概率论》一般是把《概率的哲学导论》作为其绪论考虑）。因而，《分析概率论》开创了概率论发展的新阶段，是概率论发展史上的具有”里程碑”式的著作。

1.《概率的哲学导论》的内容及拉普拉斯的概率的哲学观点

《分析概率论》第一版在 1812 年出版，分两卷，第一卷又分两册。第一卷第一册在 1812 年 3 月 23 日出版；其余部分在同年 6 月 29 日出版。在 1814 年 11 月 14 日出第二版时，增加了长达 150 页的绪论，同年单独印刷成一本书，题为《概率的哲学导论》(Essai philosophi-que sur les probabilit é s)。1820 年出的第三版中，仍保留绪论，但单独印出的《概率的哲学导论》分别在 1816 年印出第三版，1819 年出第四版，1825 年出第五版。《分析概率论》还有四个附录：附录一在 1816 年印出，附录二在 1818 年印出，附录三在 1820 年印出，附录四是拉普拉斯在 1825 年写完的，但生前未印出，以后收在他的全集第七卷中。

绪论即《概率的哲学导论》中包含概率论的发展历史及一般原理和应用，是他在师范学校的讲稿编辑而成。其中有五个内容：一是概率的定义及发展历史，拉普拉斯提出“先验”概率的概念，这是有争议的；二是概率计算的一般原理，主要是古典概率论的原理；三是讲述概率论中的一个重要概念——期望；四是讲述概率的分析计算方法，实际上是古典概率论向分析概率论过渡；五是讲概率计算的应用，占绪论的一大半，应用于各种各样的自然和社会问题。

《分析概率论》绪论的内容起源于拉普拉斯于 1795 年至 1796 年在巴黎师范学院给学生们开的关于初等数学的几次讲座。1814 年和 1820 年作为《分析概率论》的绪论出版，

从1816年开始它以题为《概率论的哲学导论》单独出版。这个小册子对概率论知识的普及推广起到了很大的作用，因为它的内容表述几乎不用任何数学符号和公式，它的目的是使广大读者不用求助高深的数学知识就能了解到概率论的基础知识，可以把这个《概率论的哲学导论》看成是他的《分析概率论》配备的一个向导。

《概率的哲学导论》共有如下内容，其目录如下：

“论概率

概率演算的一般法则

论期望

概率演算的分析方法

关于机会的游戏

存在于假设相等的几率之间的未知的不等式

从事件的无数次重复中导出的概率的定律

概率演算对于自然哲学的应用

概率演算对于道德科学的应用

关于证言的概率

关于选举和法庭判决的概率

死亡表和寿命、婚姻和一般的结合的平均持续时间

论依赖于事件概率的公共机构的福利

论在概率估计中的错觉

论达到确定性的各种方法

关于概率演算的历史评注”

上述这些内容在整体上可以分为三部分：首先是拉普拉斯对于概率论的本质，概率论的基本法则和概念，以及概率演算的一般方法的哲学理解；其次是概率论在具体领域中的一些应用的理解，如在自然科学、道德科学、观察误差的平均值、证言的可靠性分析、集体决定和选举、法庭判决、人寿保险、误差测量等几乎无处不在的应用的论述；最后一节是概率演算的历史评述。

拉普拉斯在《概率的哲学导论》中的第二章“关于概率”和第三章“概率论计算的一般原理”论述如下：

“第二章 关于概率

生活中发生的一切事件，甚至那些因无足轻重而被认为不遵从神圣的自然法规的事件，都如同太阳的公转一样是自然法规的必然产物。因为没有明白宇宙这个庞大体系中连接这些事件的每个环节，所以，它们的出现只好被视为有“终极原因”的。……但是随着人类知识的积累，那些主观臆断的所谓“终极原因”逐渐站不住脚了。而且在有说服力的哲学面前完全遭到摒弃。哲学本身只是表明我们对真正原因的无知。

现在的事件和先前发生的事件被基于“万事皆有因”。这一明显法则的纽带所连接。这一被称之为“充分理由律”的公理甚至推而广之适用于完全不在意的行为，即认为除非有确定的动机，人的自由意志是不可能的。如果有两个外部条件完全类似的场合，在一个场合下意志表现为主动，而在另一场合下，意志表现为被动，则这一意志的选择是无缘由地举动，此即莱布尼兹所谓的“伊壁鸠鲁学说信奉者的盲目的机会”。也有人认为意志是

自由的行为，是不受特定动机制约的。我们认为，存在这种看法的原因可能是没有注意到在平凡的事件中，隐蔽在意志的选择之后的基本原因。因此。我们应该把宇宙中现在的状态理解为过去状态规律地作用的结果，并且也是将来状态产生的原因。如果假设存在这样一种超常的智慧，它能了解使自然界运转的全部动力，以及构成自然界中每个物体的各自的位置，并且还能对它所知道的这些情况进行加工和分析，以致用一个公式就可把宇宙中最大的物体连同最小的原子的运动给出完整的描述。那么，在他看来，未来要发生的事跟过去已经发生的事一样，是一清二楚的。当然如果是这样的话，也就是没有什么不可确定的。但是，人类迄今为止天文学方面的知识同刚才我们所的那种情况相比，简直是可怜了。欣慰的是， 人类所具有的力学及几何学方面的知识，再加上所拥有的对宇宙引力的认识，已经是它能把自然界的运动（无论是过去的还是将来的抑或现在的）统一在为数不多的若干个具有相互联系的表达式中。更为可贵的是，还能成功地运用它们去解释在自然规律作用下发生的现象。所有这些探索真理的努力使人类的智慧逐渐地接近于上面所说的大智大慧。可是，最终达到它是不可能的。人类所独有的知识不断趋于完善是人类高于动物的一个基本原因。另外，知识的不断丰富在不同的民族、不同的时代差别尤为显著，这也就是人类文明的辉煌所在。

让我们回想在古代，一场不合时令的暴雨，一次严重的干旱，一颗拖着长尾巴的彗星，以及发生过的日食、月食、极光等一切不平常的现象都被认为是上苍惩罚人类的先兆，于是人们便祈求上苍以避免各种各样的灾难降临。然后，却没有人祈祷行星和太阳停止运行，因为观察的结果表明这种祈求纯属枉然。不过，由于这些异常的现象一般的间隔周期都要经过很长一段时间，所以不管是否它们符合自然界的规律，人们都猜测这肯定是上苍不满人间的种种罪恶才制造这些异常现象以示其不满。1456 年，就在人们对上空出现的拖着长尾巴的彗星无疑给人们又增添了恐怖的气氛。这颗彗星在出现四次以后，激发了人们的非常不同于以往的兴趣。在此之间，人们获得的关于宇宙运动情况的认识使它不再引起惊慌，这是因为惊慌只表明人类对宇宙认识的匮乏。当哈雷发现 1456 年出现的彗星和 1537 年、1607 年以及 1682 年出现的是同一颗，并且又预言它下次将会在 1758 年或 1759 年出现时，整个学术界都处在焦虑的等待中。因为预言的证实将意味着科学界中的一个很大的发现和对塞尼克预言的应验。在谈到那些从天空中坠落下来的流星时，塞尼克曾说：“可能是在经过了若干个世纪的探索，总会有那么一天，困扰人类的真相会昭然于天下，以致我们的后代会吃惊为什么真理那么容易从我们身边溜走。”此后克莱沃特开始着手木星和土星对彗星运动影响的研究，经过复杂的计算之后，终于确定了彗星要在 1759 年 4 月初在最靠近太阳的近日点上出现。这一预言最终被观察所证实。当然，在天文学中科学家们揭示的彗星运动的规律性无疑存在于其他一切现象中。

其实，单个空气或水蒸气分子的运动轨迹和行星运动的轨道一样是有规律可循的，只不过是人类缺乏对前者足够的认识。

概率部分地与这种无知有关，部分地依赖于我们的知识。倘若知道三件或更多的事件中总有一件会发生，然而又没有任何理由认为其中一件比其他事件更有可能发生，那么，在这种不确定的情况下，要说明哪一事件肯定会发生却是不可能的。但是，如果许多等可能情况都排除莫一事件发生，只有一种情况有利于它的发生，那么，在这些事件中，随意地选取一件，则该事件不大可能出现。

机会的理论就是把同类的所有事件划归为一定数量的等可能情况，这里所说的等可能性，就是对于属同一类的事件，人们以同等程度不能判定哪种事件会发生，并且确定在哪些情况下被考虑的事件可能发生。可能发生的次数与总的次数之比就是对该事件出现的概率的一个描述。

上面关于概率的概念假设了概率是恒定的，即当总的试验次数增加时，该事件发生的次数也将增加，并且比值不变。为了说明这点，假设有两个罐子A和B，A中有四个白球和二个黑球，B中有二个白球，一个黑球。假想A中有二个黑球用绳子连在一起，而当其中任何一个黑球被取到时，绳子就会自动断开，另外的四只白球同样用这种绳子连成两组。由于取到A中连在一起的两个黑球将意味着取了一只黑球，因此，若假想连接的绳子不会断开时，显然，这对取黑球的可能不产生任何影响，只不过是每回取到是两个球，因此，在两种情况下，从A中两两连在一起的三对球和B中的三个球并没有本质上的不同，由此便证实了本段开始的断言。

当某一件事件在所有试验中都发生时，可能性事件就成了必然事件，它的概率等于1。在这种情况下，可能性与确定性变得可比了，尽管它们之间有着本质的不同，例如，当人们说某件事是事实时，它就被认为是实实在在的，而当人们说某件事发生的概率为1，那就意味着对它发生的断言还有修正的可能。

在只考虑可能性事件时，掌握信息材料的多少是对同一事件可能性看法不同的主要原因。例如，假设有三只罐子A，B，C，其中一个里面全是黑球，而其余两个里都是白球，求从罐子C中抽一只黑球的概率。如果不知道哪知罐里装的是黑球，那我们就没有理由说一定是罐C而不是A或B，事实上，可能是三个罐中的任一个，此时，要求的概率为$\frac{1}{3}$。如果我们已知A中装白球，那么装着黑球的只可能是B或C，此时，要求的概率就变成了$\frac{1}{2}$。最后，如果已知A和B都装的是白球，那么，很明显，此时所求的概率就为1。

因此，关于某一件事件的认识，听众的相信度会因其知识面的宽窄而有所不同，如果对此事件报告的人的地位和品质使人能产生充分的信任感，而他本人对于所报告的内容又充满自信，那么无论他说得多么荒诞离奇，那么想得到信息的听众也会像听他平常讲话时毫不怀疑。但是，假如碰巧听众中的一位发现这人对该事件的陈述和另外一个同样值得信赖的人的陈述有出入时，那么他就会对此产生疑惑，聪明点的就会认为该事件不足为信，并且对此彻底地不予考虑，而不管是否它已被证明或悖于自然界的法则。

总有那么一些人一向被大众誉为是博古通今，而且成了生活中重大事情思考并做出判断的主宰。可是，我们说，正是因为这种情况，他们的意见才成了种种谬误的根源，而这些谬误在人类的蒙昧时代禁锢了人类思想的自由发展，古时候常提出的魔法和炼丹术即是两个恰当的例子。这些错误思想在幼年时就被灌在头脑中，并且要求可以不加验证地采纳。大家都说它对，那么它就是对的，这种错误的思想持续了相当相当长一段时间。随着社会科学的进步，一部分具有远见卓识的人开始认识到这些谬误，而他们的观点通过模仿和习惯广泛传播，使那些谬误现在在一般人的心目中已不再存在了。这是精神世界里取之不竭的力量源泉，它可以在其中建立与原来具有同样信誉然而观点却截然不同的思想。所以，

如果我们知道了看法的差异往往囿于种种偏见时，我们就会没有理由不去关注与己不同的看法。因此，只有对自己观点和主张做出认真的检验和公正的评价之后，我们才能去教导那些被我们认为知识贫乏的人。

一般来说，不同的意见往往产生于对已有的信息的理解和使用上。在概率论中往往考虑一些如此微妙的问题，以致由同样的信息得到不同的结果是非常平常的，特别，对于那些复杂的问题尤为如此。下面考虑概率论中的一般原理。

第三章 概率论的一般原理

原理一：原理一是概率的定义，如前所述，它是有利情况的个数与所有可能情况个数之比。

原理二：假定所有情形都是等可能的，否则，则要先确定它们各自的概率（它们的精确估计正是机会理论中的一个难点所在），然后将每种有利情况的概率相加就是所求的概率。下面用一个例子来说明这个原理。

假设我们向空中抛一枚大且很薄的硬币，硬币的两面分别称为正面和反面，并且假定这两面完全相似，现在来考虑抛两次，硬币落地时至少有一次正面朝上的概率，显然有四种等可能情况发生：前两次都是正面；第一次正面，第二次反面；第一次反面，第二次正面；两次都是反面，明显地，前三种情况符合要求。因此所求的概率为$\frac{3}{4}$。如果打赌的话，抛两次至少一次正面朝上的胜算为三比一。

在上面的游戏中，也可以把情况分作三类：第一次正面，而第二次可以是正面也可以是正面也可以是反面；第一次反面，第二次正面；最后一种情形，两次均为反面。按照达朗贝尔的说法，上面三种情况是等可能的，这样，要求的概率就成了$\frac{2}{3}$。但是，显然第一次出现正面的概率是$\frac{1}{2}$和第一次反面、而其他两种情况的概率分别为$\frac{1}{4}$，第一种情况作为一种简单事件，实际上是两个事件的复合：第一次正面、第二次正面以及第一次正面、第二次反面。如果按照原理二，把第一次正面的概率$\frac{1}{2}$和第一次反面、第二次正面的概率$\frac{1}{4}$相加，得到所求的概率为$\frac{3}{4}$，这和前面对它做两次抛处理所得结果一样。由此可见，先前的做两次抛分四种情况处理并未改变打赌者取胜的机会。这种做法目的仅仅在于使得不同的情况等可能地出现。

原理三：这里概率论中最重要的原理。概率论中最令人迷惑不解的是不同事件组合在一起，它们同时发生的概率为何时而增大，时而又减小。如果几个事件是相互独立的，那么它们同时发生的概率就是它们单个发生时概率的乘积。因一掷骰子掷出一点的概率为$\frac{1}{6}$，掷两次掷出两个一点的概率就是$\frac{1}{36}$。掷两次中，其中一掷的任何面都可以和另一掷的六个面相组合成一种情况，所以，两掷共有 3 6 种情形，但是，其中只有一种是两次掷出一点。一般来说，相同条件下，某事件连续重复若干次概率等于该事件出现一次的概率的方幂，方幂指数等于重复的次数。由于一个真分数的方幂增大时，其值将不断地减小。

因此，一种可能性非常大的事件，在多次重复的过程中连续发生就变得十分不可能。假想，某一事故由二十名目击者以你传我、我传他这样的方式依次传下去。并且假设每一次传递的可信度为$\frac{9}{10}$，则经过二十次后，可信度将不及$\frac{1}{8}$。关于这样的可能性的衰减，有一个非常形象的比喻：物体发出的光在经过许多次镜面反射后就消失殆尽了。即使是为数很少的几面镜子都可以给我们一个清晰的图像。历史学家对这种经过许多年代之后，事件可靠性降低却很少给于足够的重视。易于这种事实，很多一直用最肯定的语气来描述的历史事件至少应该成为疑问。

在纯数学领域中，只要原理正确，那么由此得到的最间接的结论也被认为是正确的。例如微积分运用于物理学而得到的结论与事实或实验是完全相符的。但是在理论学中，每个推断仅仅是以一种大概的方式从它的前提推出的。所以，无论这种推理看起来多么合理，错误产生的可能还是随着这种推理次数增加而增加。如此，最后得到的结论也就很可能不正确。

原理四：如果两个事件彼此相关，那么它们同时发生的概率等于第一个事件发生的概率乘以在第一个事件已发生的前提下，第二个事件发生的概率。因此，在前述的三个罐子A，B，C里（其中两个里装的白球，另一个里面装的黑球），从C里抽取白球的概率是$\frac{2}{3}$，假设如果已经从C中抽到了一只白球，那么只需在A、B中间判断那个里面装了黑球。而从B中抽到白球的的概率为$\frac{1}{2}$，此时乘积$\frac{2}{3}\times\frac{1}{2}$，即$\frac{1}{3}$，就是同时从$B$，$C$抽两只白球的概率。

从这个例子中可以看到过去事件对将来事件概率的影响。最开始从B中取一只白球的概率为$\frac{2}{3}$；当已知从C中抽取一只白球时，所求概率就变成为$\frac{1}{3}$；而当已知从B中抽取出一只黑球时，可能事件就变成了确定事件，相应的概率为0。现在根据以下的原理来确定这种影响，这一原理可以看作前一原理的推论。

原理五：如果由该事件和另一待考察事件复合而成的事件的概率，那么第二个概率除以第一个概率之后就是待考察事件的概率。

这里的计算法涉及某些哲学家提出的已发生事件对将来事件发生的概率影响的问题。假如在抛硬币的游戏中，正面出现的次数总大于反面出现的次数，那么我们就要猜想是否在硬币的构造中存在某种原因是硬币正面朝上的机会增多，这正像在生活中，老板倾向于雇佣随和的人一样。可是若是环境上是不可信赖的话，那我们就又会回到原来那种不知所措的情形。例如，如果每抛一枚硬币就换一枚，这样一来，过去的假设对将来不会产生任何影响，所以如果在考虑过去将被认为是荒唐透顶。

原理六：如果导致某一被考虑事件的发生有各种原因，每种原因的可能性设想为导致该事件发生的概率，(当然，这里必须保证所考察的始终是同一事件。)某种原因成立的概率是个分数，分子是该原因下该事件发生的概率，分母是相应原因下该事件发生的概率总和。如果先验地考录一些非等可能的原因时，那么在计算分母时，不用该事件在每个原因下发生的概率，而必须用这一概率乘以该原因自身的可能性。这就是所谓的关于从事件过

渡到原因这一机会的分析分支中的最基本原理。

这一原理解释了为什么我们认为有规律的事件时有特定的原因所致。某些哲学家认为规律事件的发生远较其他事件发生的可能性要小。例如，在抛硬币中，连抛二十次均为正面本质上比正面、反面随即出现要难得多。然而，这种观察先假定了过去事件会影响将来事件发生的可能性，而这点确实令人无法接受。规律事件出现极少是因为这类事件的数目本来就少。当发现了事物的某种对称性，并为此寻找解释时，与其认为对称性事件比其他事件来的少，不如说它是一个有规律的原因起作用的结果，要么就说纯属于偶然。(其实，多数情况下，我们更倾向于认为是前者)例如在一张桌子上，我们看到字母的如下排列“C onstoantinople”，则会认为这种排列决非偶然，这并非因为这种排列比其他排列出现的可能性要小。原因在于：如果这个词不出现在任何一种语言里，那么人们将不会贸然断言它可能出于故意。问题是这个词是在我们当中被使用的，这样一来，上述的排列非常可能是出于有意，而不是偶然。

下面来解释所谓的“离奇”。在人们的心目中，将所有的事件分为几类，包括事件很少的那类就被认为是“离奇”的一类。因此在抛硬币中，正面连续出现 100 次则被认为是离奇的。因为在连续 100 次的抛硬币中……”

《概率论的哲学导论》是拉普拉斯本人关于概率论这一学科及各部分具体内容的哲学理解的总结。这些哲学思想是他的概率理论研究的基本出发点，并且直接导致了他对概率论这一学科的认识，进而影响了他在概率论实践中所采用的方法。

18 世纪以来，随着牛顿力学所取得的辉煌成果，以牛顿为代表的一种机械决定论的思想在欧洲科学界蔓延，在这样一种精神的氛围中，决定论思想在拉普拉斯那里发展到高峰。孔多塞、Boscovich 和 Maupertuis 等人都是牛顿思想的继承者和崇拜者，他们对拉普拉斯的决定论思想的形成具有直接的影响。在第一节“论概率”的开头，拉普拉斯就提出来在科学史上著名的一段话，这段话被看做是拉普拉斯的决定论思想的最明确的表白：

“我们应该把宇宙的目前状态看作是它的先前状态的结果，并且是以后状态的原因。万能的指挥者能够在某一瞬间理解使自然界生机盎然的全部自然力，而且能够理解构成自然的存在的各自的状态，如果这个智慧者广大无边到足以将所有这些资料加以分析，将宇宙中最巨大天体的运动和最轻的原子的运动都包含在一个公式中。那么对于这个智慧者来说没有任何事物是不确定的，未来如同过去一样在他的眼中将一览无余。人类的心志在致力于天文学上所表现出来的完美中，给出了与这一智慧者微弱的相似性。人类在力学和几何学上的发现，加上万有引力的发现，使它能够用相同的分析表达式去理解宇宙系统的过去状态和未来状态。通过把同一方法应用于某些其他的知识对象，人类已能将观察到的现象归结为一般规律，并且预见到给定条件下应当产生的结果。所有这些探索真理的努力都倾向于引导人类心智不断地回到我们刚才所提到的广大无边的智慧者那里去，但是，人的心智距这一智慧者仍有无穷之远”。

既然如此，这个万能的智慧者并不需要概率论的知识，然而，对于人类来说，概率是一个不可或缺的知识，“概率一方面是由于我们的知识，一方面又是我们的无知”。概率是对人类无知的一个重要的补偿。机会对于拉普拉斯来说并不是不可化约的随机，而是一大群独立事件的相互交错和相互作用这样一个图式的偶然的结果。如果自然界所有事件能够同时被感知到，并且如果我们的演算技术足够先进的话，那么我们将不比无限的智慧者

更需要概率。但是有限的人类是不可能达到万能的境界的，这样概率本质上是对人类错误的水平的一种估计，概率应用于自然界也只是人类在一定的知识水平上的预测，而它的本质恰恰是人们可怜的无知。

拉普拉斯认为，甚至一个非常小的事件的出现也不会不遵守伟大的自然规则，就像太阳的旋转那样。但是由于我们的无知，我们不得不把它们的发生归于偶然。没有任何事情会毫无缘由地发生，即使自由的意志，不管多么的“自由”，也不会没有导致行动的动机。此时自然科学取得的如此辉煌的胜利给人类提供了一个典型的样板，那么同样地通过拉普拉斯以及同时代的其他人所设想的一种数学方法也可以应用于道德科学。拉普拉斯把关于人类的科学视为一种社会机械学，人类的可呈现的每一种情感和智力的现象几乎都可以比喻为物理中的现象。例如他说人类社会中的突兀变化分散了“活动的社会当量”，并且把激起的共同的情感反映状况比喻为“情感的共振”。尽管他承认道德的原因比对它们的物理分析更为繁杂，但是拉普拉斯还是倾向于这一点：他的机械的和动力学的比喻远比单纯的比喻公正：“在两种对立的动机之间的犹豫不决就是等作用力的一种平衡……，过于的专注，或者说连续不断的努力会使神经中枢疲惫不堪，就像长时间的连续的点击会损耗电堆和鱼的器官一样。为了呈现可感觉的智力的事情，我们从物质的对象提炼出的几乎所有的比较在本质上是相同的”。

拉普拉斯的社会机械学的思想还体现在社会规律的稳定性方面，就像物理系统的均衡状态一样。统计数据的稳定性是这种思想的最明显的体现。例如，拉普拉斯说从法国的彩票中得到的收入是稳定的。

拉普拉斯还认为探索宇宙的过去状态与探索宇宙的未来状态最终可以划归为概率论中的两大类问题：一类是由事件的结果探索导致事件发生的各种原因的概率，另一类是已知原因探讨将要发生的事件的概率。这两类问题是概率论研究的最主要的目标。机会的游戏中的许多问题为这两类问题提供了典型的数学模型，例如，摸球的模型。但是何以达到这个目标？有两个工具为达到这个目标提供了可能，这两个工具就是伯努利定理和贝叶斯定理。根据伯努利定理，对现实的完整的观察和记录会逐渐地为可能发生的事件提出好的猜想，反过来，由贝叶斯定理根据可能的证据去检验已发生事件的原因。总之，探索宇宙的方法通过伯努利定理和贝叶斯定理一前一后的合作，再加上日趋完善是数学分析技术，人类会逐渐地、一步一步地向真理接近，人类的认识会由统计数据所逐渐地完善。

在拉普拉斯的思想深处，其最主要的一个目标就是揭开揭示原因的定律。这项工作占据了他的概率研究的大部分。拉普拉斯把自然界看做是“规则的”与“不规则的”原因的复合体。但是在长期的趋势中不规则的原因的影响是对称的，这些影响彼此抵消，这一点暴露了自然界是由恒定的原因所操纵的。恒定必然胜利，规则因最终可以征服不稳定的因素。除了统计稳定性的无数事实，概率论中的许多定理也证明了这一点，例如，中心极限定理。在拉普拉斯的概率工作中，他对于摸球问题相当重视，在这点上他继承了伯努利的观点，认为许多自然和社会的规律可以由这个罐子模型来解释，在他的《概率的哲学导论》中他曾这样用摸球模型演示中心极限定理：

“想象把一系列罐子排成一个圈，其中每个罐子中有大数目的球：白球和黑球两种，起初白球和黑球的比在这些罐子中是不同的。例如，一个罐子可能仅有白球，而另一个罐子中可能仅有黑球。从第一个罐子中抽取一个球放入第二个罐子中，将第二个罐子的球搅

拌均匀，再从第二个罐子中抽出一球放入第三个罐子中，这个过程持续下去，直到从最后一个罐子中抽取一个球放入第一个罐子中。然后这个过程重复地一次一次进行下去。概率的分析向我们显示，在这些罐子中白球和黑球的比将以等同而结束，并且等于白球的总数和与黑球的总数的比。这样根据这个变化的规则图式，在这些比之间的初始的差别随着时间而消失，而让位于简单的秩序。现在，假设在原来的罐子之间放一个新罐子，并且新罐子中白球与黑球的总数与原来罐子中的黑球与白球总数不同。如果在混合的罐子中重复地一次一次进行我们刚才讨论的程序，那么在原来罐子中建立的简单秩序将被打破，且白球与黑球的比将从一个到另一个有相当的差异。但是这种差异将一点点消失，最终让位于新的秩序——罐子中白球与黑球有相同的比。这些结果或许可以推广到自然地正在发生的组合，其中，给元素以活力的永恒不变的力量建立了行为的规则图式，从而揭开了隐藏在一片混沌的迷雾中的由这些令人敬畏的规律所统治的系统体系”。

所以对拉普拉斯来讲，概率论归根结底是一个揭示和发现隐藏在缤纷杂乱的现象之下的永恒规律和法则这一伟大目标的工具，它的主要价值就是帮助人类实现这一目标。在拉普拉斯的《分析概率论》这部内容繁杂、头绪众多的著述中，其形式结构在表面上给人一种松散无序、迷宫一般的感觉，但是其内容却始终贯穿着一个意图：概率是人类揭示自然界和人类社会中那些令人敬畏的规律所必需的一个工具。这种意图把这些纷繁的内容连接成前后一致的整体。所以在《概率的哲学导论》中的大部分的篇幅中以及《分析概率论》的第二卷中，在不同的层次上详细地讨论概率论几乎无所不在的应用等就是拉普拉斯的决定论思想的体现。所以决定论的思想并非像有些学者所理解的那样是阻碍了概率论的发展，其实，它是拉普拉斯及其继承者们以极大的热情投入其中的强大动力和出发点之一，拉普拉斯的概率论在所有领域中的实践几乎都是建立在一个假设上的，即所有的现象是稳定的和有秩序的，由此使得“理性的人”对于所有现象的量化和演算是可能的。

2.《分析概率论》第一卷的内容简介

第一卷有标题为：“生成函数的计算” (Calcul des fonctiong é n è ratrices)，主要讲述同概率计算有关的数学方法，共分两册五章。上册讲述带有整数指标的函数族，当指标数量很大时的一般情况。这是因为概率讨论中，要用到试验次数很大，重复次数很多的现象。所用到的特殊函数同这些次数有关。内容只有两章。第一章讲一个变元函数族的情况，第二章讲两个变元情况。这些知识都是他本人在二、三十年前有关级数理论的研究成果，本书中仅为系统化而已。下册是讲述大指标函数的近似理论，这些特殊函数是生成函数的级数展开式系数，一般要满足某种微分方程或差分方程，要用各种近似方法积出它们。下册共分三章(即第三、四、五章)。第三章讲非线性微分方程的近似积分方法；第四章讲线性差分和微分方程的近似积分方法；第五章讲前两章的方法用于求出大指标函数的近似。整个第一卷是为第二卷作的数学准备。

第一卷(卷 I)题为“生成函数的计算”这一部分的内容几乎是拉普拉斯在 18 世纪 80 年代的两种重要的数学探索的再版，当然有些修订。1872 年的那篇题为“论级数”的专门论述生成函数的论文成为卷 I 的第一部分的基础。1785 年和 1786 年两篇论述包含大数的定积分式子的近似求法的文章成为卷 I 的第二部分的基础。而卷 I 的引言引述了他的 1811 年的论文中的引言，大意是两种理论是一个学科的两个分支，一个是有关用之于表达随机事件的微分方程的解，另一个是当事件被重复许多次时估计哪个结果的表述更精确。拉普

拉斯在引言中还说他现在要以比三十年前更一般的方式展示这些理论。在法则上，主要的不同是应用了关于指数量的分析处理的最新发展的计算技术。在一开始的历史回顾一节中，拉普拉斯追述了一条从拉格朗日、莱布尼兹、牛顿和沃里斯等人的工作一直到笛卡尔为了进行平方、立方和提高的整数次幂的运算而发明的指数的一条历史发展的线路。

在卷Ⅰ的开头，他给出了与他的早期论文中的开头一样的生成函数的定义：

"设y_x是x的任意函数，如果可以形成这样一个无穷级数

$$y_0+y_1t+y_2t^2+y_3t^3+\cdots+y_xt^x+y_{x+1}t^{x+1}+\cdots+y_\infty t^\infty,$$

设u表示这个这样的级数的和，或者其函数的展开式形成这样的级数，那么这个函数就是我所称的变量y_x的生成函数"。

他进一步解释到：一个变量y_x的生成函数式一个可以被展开为t的幂级数，且拥有t^x的系数的变量，反过来一个生成函数所对应的变量是函数以t的幂的形式展开式中t^x的系数，t的幂的指数式中就表示了变量y_x在级数中所在的位置，其中也可以根据负指数的形式无限地向左延伸。

实际上两篇早期的论文和卷Ⅰ的修订之间的主要区别式拉普拉斯省略了某些繁杂的段落，并给予其他关键性的内容以突出的地位。例如，他省略了论文"论级数"中的关于二阶偏微分方程的解，这个问题对于物理中的问题非常重要，但是对于机会游戏的理论就不那么重要了。另一方面，拉普拉斯却更重视早期论文中对于从有限到无穷小，从实数到虚数有关段落的强调。他赋予在实数量过渡到虚数量的过程中不断持续发现的丰富多彩的结果以极大的重要性，并在从第一部分过渡到第二部分的段落中专门讨论这些被转化成收敛级数的定积分的极限由一个方程的根所给出的过程步骤，在这个推演中，当系数的符号变化时，根就变成了虚数。这个性质引导拉普拉斯对于经常出现在概率研究中的以及依赖于两个超越数π和e的某些定积分的值进行研究。与他早期的适用面窄而又不直接的方法相比，这里给出的是一个估计这些积分的完善的直接方法。

以上是卷Ⅰ的大概结构和内容，这些方法都是属于纯数学的方法，关于概率论本身的探讨在这一卷中一点都没有涉及。这一部分主要限于生成函数的计算的数学方法和它在一般的纯理论中的应用。例如，在插值法和级数的变换、微分方程的求解和用定积分的术语表述函数等。拉普拉斯的主要目的是想把生成函数作为概率论的基础，就像他在1811年的论文中所说："生成函数的演算是我打算不久就要出版的关于概率的一个理论的基础"。拉普拉斯没有食言，在他的卷Ⅱ中生成函数和分析学的方法渗透在他对于概率论中大部分问题的解决中。

3.《分析概率论》第二卷的内容简介

第二卷的标题为"一般概率论"（Théoré générale des probabi-lités），是本书的主要内容，共约400页。整卷分为十一章，全面归纳了前人和他自己有关概率论的成果，并应用于自然哲学、天文学、大地测量学、测试、误差、审判过程和选举机构等方面的问题。不像《天体力学》前两卷那样系统化讲述。虽不宜作为教材，但作为概率论的研究参考资料是很好的。各章自成系统讨论某一方面问题，但相互之间独立性较强。

拉普拉斯对于各章的顺序安排主要是由于实际的应用决定的，它们之间并非相互独立，如果从更广泛的角度来分，这些章的内容可以分为三大部分：第一部分是现实意义上

的概率论本身，主要包括一般概念的引进，人口统计中一系列重要问题的解决，概率演算中怎样应用有限偏微分方程的方法，在自然和社会统计中没有直接应用的一些整数序列问题的解，道德期望的研究和证言的可信度和证据的正确性的估计(第一、二和八至十一章)；第二部分是极限理论(第三、五、八、九章)；第三部分则是数理统计(第四至八章)。在卷Ⅱ中也不时重复穿插拉普拉斯对于这些课题的哲学解释。

第一章：理论的一般法则

第一章讲概率的一般原理和学科特点。除了提出概率的古典定义和概率的乘积原理外，作为第三个基本原理是文字叙述的一个定理。这是有关由 n 种不同原因产生的事件的概率公式的定理，即在 1763 年已发表的贝叶斯定理。可是，拉普拉斯在书中没有提到贝叶斯(Bayes)的名字，而作为自己的一个定理。接下去的几章讨论各种实际问题中提出的概率计算方法。

本章大多是对《概率的哲学导论》的前几部分的重复并给予数学化的处理。在这一章的开始，拉普拉斯首先给出我们现在所称的事件的概率的古典定义：“事件的概率为有利事件总数与总的事件总数之比。当没有什么使我们相信一个事件比其他事件更应该发生时，这些事件对我们来讲就是等可能的”。这个概率的定义最初的形式的萌芽由卡尔达诺在其《游戏机会的学说》中引进，但是，对早期概率的定义的发展产生最大影响的是莱布尼兹的工作。在一个题为标注的日期是 1687 年 9 月的备忘录中，莱布尼兹宣称“概率是可能性的度量”。莱布尼兹长期以来一直把等可能性与概率联系起来，尤其是把有关概率的论述建立在一个后来被称为“中立的法则“之上，这个名称首先是由凯恩斯在他的《论概率》中给出的。这个法则还有一个名字为“不充足理由律”，是由克莱斯在 1871 年的一本教科书中首先采用的。法则的内容是：“当没有什么使我们相信一个事件比其他任何事件更应该发生时，这些事件对我们来讲就是等可能的。”哈肯认为莱布尼兹的概率定义是拉普拉斯的概率定义的最终源泉，受莱布尼兹的影响，伯努利和棣莫弗两人都明确地提出过这个定义，他们的著作对拉普拉斯关于概率的定义有直接的影响。对拉普拉斯产生最直接影响的还有一个人就是孔多塞。在 1785 年把概率论应用于投票理论的一书中，孔多塞给出的概率定义具有“纯粹数学的意味”：概率是等可能的组合之比。之所以说拉普拉斯也受到孔多塞的直接影响是因为拉普拉斯在 1774 年和 1776 年的论文中提出概率的古典定义时还没有如此完整和明确，因为那时孔多塞的工作还没有发表。在《分析概率论》的第一章中拉普拉斯将伯努利和棣莫弗的定义与孔多塞的定义结合在一起。所以从莱布尼兹、J.伯努利、棣莫弗、孔多塞到拉普拉斯的古典概率的定义具有一脉相承的关系。

在此卷接下来的大部分篇幅中，拉普拉斯阐述和证明了不相容事件(即不能同时发生的事件)的加法定理

$$P(A \cup B)=P(A)+P(B)$$

他还证明了独立事件(即其中之一的发生不影响另外一个事件的概率)乘法定理

$$P(AB)=P(A)P(B)$$

若事件不相互独立，他引进了条件概率表示乘法的运算

$$P(AB)=P(B|A)P(A),$$

$$P(A_i|B)=P(B|A_i)/\sum P(B|A_i),$$

其中 A_i 是等可能的和互不相容的事件，即 $P(A_i)$ 为常数，B 是观察到的事件。

上述公式已经推广到包括$P(A_i)$不相等的情况，这是拉普拉斯在这个问题上超出伯努利和棣莫弗的工作之处。拉普拉斯在这里实际上已明确地用到了全概率公式

$$P(B)=\sum P(B|A_i)P(A_i)$$

关于这个问题拉普拉斯还问到，假设条件只是近似的满足，从上述接受的条件中所引起的误差是多少?他用掷一个不规则的硬币的模型来讨论这个问题。最后拉普拉斯引进了数学和道德期望的定义，关于拉普拉斯的期望的论述将在第十章有较详细的论述。

值得注意的是除了道德期望和逆概率公式，这一章的其他内容在棣莫弗的《机会的理论》中也出现过。

第二章：由概率已知的简单事件复合而成的事件的概率

第二章讨论由已知概率的简单事件组成的复合事件的概率计算，此问题讨论得非常详细，几乎占了第二卷的四分之一。他从古典的抽彩、摸球等问题出发，归纳为较一般的数学问题：袋中装有$n+1$个球，各球号码依次为0，1，2，…，n；每次摸出一个，记下号码后再放回袋中；则摸出i次后，i个球的号码总和为s的概率是多少?

设每次摸出球的号码分别为$t_1,t_2,\cdots,t_i$，显然有

$$t_1+t_2+\cdots+t_i=s \quad (2-1)$$

拉普拉斯推出这个概率公式为

$$\frac{1}{n(i-1)!}[C_{s+i-1}^{i-1}-C_i^1C_{s+i-n-2}^{i-1}+C_i^2C_{s+i-2n-3}^{i-1}-\cdots] \quad (2-2)$$

其中组合符号定义为

$$C_n^r=\frac{n!}{r!(n-r)!}=\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}(r!=1\cdot2\cdots r)$$

式2-2的级数到$(s-n),(s-2n-1),(s-3n-2),\cdots$为零或为负数时为止。若$s,n$无限增加时，式2-2可化为

$$\frac{1}{n(i-1)}[(\frac{s}{n})^{i-1}-C_i^1(\frac{s}{n}-1)^{i-1}+C_i^2(\frac{s}{n}-2)^{i-1}-\cdots] \quad (2-3)$$

此式后用来讨论彗星轨道倾角分布于$\Phi-\varepsilon,\Phi+\varepsilon$之间的概率，$\Phi,\varepsilon$为任给定的角度，但$\varepsilon$为小量。

在本章中拉普拉斯以一个数的祝贺公式的初级解释作为开始，继续讨论了一系列的求直接概率的问题：古典的抽彩的问题、摸球问题、赌徒输光问题等。这些问题讨论得非常详细，几乎占了卷Ⅱ的四分之一的篇幅。

(1)在一个给定的抽取中，抽取彩票的所有数目至少一次的概率。

每次抽取相当于在n个数中抽取r个数。在i次有放回的抽取中所有数目至少被抽到一次的概率是什么?这个问题也曾出现在拉普拉斯以前的论文中，一个等价的问题也曾由棣莫弗解决过。不过拉普拉斯在这里用有限微分方程形式转化了这个问题，这个方程的解需要非常复杂的变换。同样的问题也由后来的欧拉于1781年解决过。

(2)从一个罐中抽取若干个标号是偶数的球，即一个罐中有x个球，从中抽出几个。

假设抽取标号为n的球($n=1,2,3,\cdots,x-1,x$)的等可能是相等的，求抽取一个标号是偶数的概率。这类问题具有悠久的历史，在早期，简单的奇数和偶数的游戏是出于占卜和预

测的目的。这个问题的数学研究至少可以追述到 1730 年一个名叫 Mairan 人的工作中。拉普拉斯本人在此之前曾两次解决过这个问题。不过在他的早期的论文中，他是用有限差分方程的方法解决了这个问题。

在这里，拉普拉斯将解法更加简化，这个解法是在以下结论的基础上给出的。

可能性的总数是

$$(1+x)^x-1,$$

其中抽取标号为奇数的球的有利数是$\frac{1}{2}(1+1)^x$个，抽取标号为奇数个球的有利数为$\frac{1}{2}[(1+1)^x-1]$个。

拉普拉斯还进一步把这个问题引申：一个罐子包括X个白球X个黑球，并且抽出偶数个球，求抽出相同数目的白球与黑球的概率。此问题有利的可能性的数目(有利事件数)是

$$(C_x^1)^2+(C_x^2)^2+\cdots+(C_x^x)^2,$$

拉普拉斯注意到在这个和中再加上一项$(C_x^0)^2$，就等于下面的乘积

$$(1+a)^x(1+\frac{1}{a})^x=\sum_{k=0}^{x}C_x^k a^k\sum_{k=0}^{x}C_x^k=\frac{(1+a)^{2x}}{a^x}=C_{2x}^x,$$

这是一个独立于a的项。

另一方面，总的可能性的数目是

$$\frac{1}{2}[(1+1)^{2x}-1],$$

所以，所求的概率是

$$\frac{C_{2x}^x-1}{\frac{1}{2}[(1+1)^{2x}-1]}.$$

(3)从几个罐子中抽取给定数目的一种或另一种颜色的球的概率。

设第一个罐子中有p个白球，q个黑球，第二个罐子中有p_1个白球，q_1个黑球，……。共有$x+x'$个罐子，如果从每一个罐子中依次抽一球，共抽取$x+x'$个球，对抽取x个白球和x'个黑球的有利可能性数等于

$$(p+q)(p'+q')(p''+q'')\cdots$$

的展开式的每一项的和，其中$p,p',p'',\cdots$的指数和等于x。

这个问题可以特殊化为$p=p'=p''=\cdots$，$q=q'=q''=\cdots$的情形，也可以推广为三种颜色的球或更多颜色的球。

(4)依特定的顺序抽取不同颜色球的概率。

有几个罐子包含p个白球，q个黑球，r个红球……，问在抽取x'个黑球之前抽取x个白球的概率是多少?或者问在抽取x''个黑球之前抽取x个白球的概率是多少?……。

就像拉普拉斯本人在解决第三个问题时，把问题限制在一个罐的情形并且在每次抽取后再用球替换。在这儿同样如此，所求的概率为

$$\sum_{(x-1)!f!f'!}^{x+f+f'+\cdots}p'^x q'^f r'^{f'}\cdots,$$

其中 $p'=\frac{p}{p+q+r+\cdots}$, $q'=\frac{q}{p+q+r+\cdots}$

这个结果包含多种情况，可令 $f,f',\cdots$ 从 0 到 $f'=x'-1, f''=x''-1,\cdots$

拉普拉斯曾多次解决了同样的问题，这次是通过一个有限偏微分方程的方法。他注意到这个问题与著名的“点数问题”有直接的联系，并且对两个各自缺乏 x 和 x' 点的参赌者来说，他们赢得赌注的概率各自为

$$p'^{x}[1+xq'+\frac{x(x+1)}{1\cdot 2}q'^{2}+\cdots+\frac{x(x+1)\cdots(x+x'-2)}{(x'-1)!}q'^{x'-1}],$$

$$q'^{x}[1+x'p'+\frac{x'(x'+1)}{1\cdot 2}p'^{2}+\cdots+\frac{x'(x'+1)\cdots(x+x'-2)}{(x'-1)!}p'^{x'-1}],$$

这两个是式子的和是 $(p'+q')^{x+x'-1}=1$。

然后拉普拉斯又考虑了赢回各自的那一份的概率的特殊情况，他得出了一个对应的有限偏微分方程。但是在解这个方程中，拉普拉斯的解法并不完全正确，这个错误他后来意识到了。

(5)以特定的顺序抽出带有标号的球：一个罐中包含 rn 个球，其中 r 个球的标号是 1，r 个球的标号是 2，……，r 个球的标号是 n，一个接一个地抽出，而不替换。所求的是在以下各自条件下的概率：

①至少有一个球在对应的位置上被抽出，例如，标号为 s 的球在第 s 次被抽到。拉普拉斯用一个纯祝贺的方法得到下面的公式

$$\frac{1}{A_{rn}^{n}}[nrA_{rn-1}^{n-1}-\frac{n(n-1)}{1\cdot 2}r^{2}A_{rn-2}^{n-2}+\frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}r^{3}A_{rn-3}^{n-3}-\cdots],$$

其中 A_{rn}^{n} 表示来自于 rn 个元素中抽取 n 个元素的排列数。

②至少有 s 个球在它们对应的位置上被抽到。在这种条件下，解法与条件①下是相似的。

最后拉普拉斯还近似地计算了没有一个球在它所对应位置上被抽到的概率。然后，他就尝试着将问题一般化为 i 个罐子的情形，但是，拉普拉斯在这里只给出解法的轮廓，而没有解法的细节。

上述这个问题等价于遭遇问题，兰伯特等人对这个问题反复思考过。拉普拉斯在这里的创新之处在于其解的过程中有两个数学表达式是他最先推出的。

(6)赌徒输光问题的概率。

所谓赌徒输光问题也称为游戏持续事件问题，它的内容是，赌徒 A 有 a 个筹码，赌徒 B 有 b 个筹码，并且用 p 和 q 分别表示他们赢得每一局的概率，问 B 在不超过一个给定的局数中输光的概率是多少?这个问题在概率论早期的历史中具有较大的影响，从现有的资料看，最早提出这个问题是帕斯卡在 1656 年与费马的通信中，除了著名的”点数问题”之外，帕斯卡还提出了其他的问题，赌徒输光问题就是其中之一。帕斯卡给出了解决方法，费马也用自己的方法加以解决，其后惠更斯在他的 1657 年的书中又提出了与帕斯卡的问题等价，与现在的提法一致的问题，并且给出“Gamble’s Ruin”(赌徒输光)的名字。他考虑的是在不考虑局数的情况下赌徒输光的问题，惠更斯仅仅给出了答案。第一个发表的解是由 J.伯努利给出的，后来棣莫弗、拉格朗日等人也对这个问题教学过研究，这个问题

逐渐演变成一个更有意义的”赌博持续时间问题”，赌徒输光问题现在被看做是一维随机过程的一般问题之一，随机过程是现代概率论的最重要的一个内容之一。

拉普拉斯在他的早期论文中曾两次考虑这个问题，那时他用的都是极其复杂的有限偏微分方程的方法。在 1810 年的一篇论定积分的论文中，拉普拉斯对$a = p, a = \infty$和一个不定的有限n，得到一个关于所求概率的方程

$$y_b = \frac{1}{2}y_{b+1} + \frac{1}{2}y_{b-1},$$

这个方程的解是

$$y_b = \alpha + \beta b,$$

他注意到$\beta = 0$和$\alpha = 1$，这样$y_b = 1$。

另外，对于赌徒持续的局数不超过n的概率y_{bn}，有一个表示这个概率的有限偏微分方程

$$y_{bn} = p y_{b+1n-q} + q y_{b-1n-1} (p = q = \frac{1}{2}) \tag{2–4}$$

其中，如果$n < b$，则$y_{bn} = 0$；如果$b = 0$，则$y_{bn} = 1$成立。拉普拉斯突然地给出这样一个式子

$$y_{bn} = 1 - \frac{2}{\pi}\int_0^{\pi/2} \frac{\sin b\varphi (\cos\varphi)^{b+2i+1}}{\sin\varphi} d\varphi, 2i = n - b$$

他计算出了上式的积分，并发现了对于赌徒B输光所需要的可能盘数。对于$b = 100$，证明了在$n = 23780$或 23781 之后，赌徒B输光的概率是$y = 0.5$。

在上式中，拉普拉斯计算积分的方法包含把被积函数展开成幂级数的过程，假设$b + 2i$很大，即n很大，他得到

$$\int_0^{\pi/2} \frac{\sin b\varphi (\cos\varphi)^{b+2i+1}}{\sin\varphi} d\varphi = \int \frac{1 - \frac{a^2}{6}\varphi^4}{\varphi} \sin b\varphi e^{-a^2\varphi^2} d\varphi, \tag{2–5}$$

其中$a^2 = \dfrac{b + 2i + \frac{2}{3}}{2}$。

然后，拉普拉斯假设，如果不引进一个适当的误差，积分的上限有可能是无限的，最后，他计算了这个关于b的不定积分的导数，这个结果为他计算积分本身提供了条件。

在《分析概率论》中，对于比较小一些n的情形下，拉普拉斯通过生成函数$p + q$的方法，这种方法棣莫弗在其《机遇论》中就已解决了。然后，他利用方程 2–4，对于$p \neq q$的一般情况，给出了一个极其困难的解法，而对于$p = q$的情况，拉普拉斯将它最终化归为表达式 2—5。

(7)赢得一个随机游戏的概率。

在对机会性游戏的研究中，一个游戏如果由两个玩家参加的情况已被反复研究过，现在考虑由$n + 1$个玩家参加的情形。一个参加者输了给定的盘数就被淘汰，而赢者与下一个参加者继续玩，游戏一直进行导有一个人连续赢n盘为止。

拉普拉斯之前的几个学者已研究过这个游戏问题，而拉普拉斯在这里的新颖之处是利

用了一个相对简单的有限差分方程的解法。其解法的思路是

令 z_x 为游戏在 x 盘后正好结束的概率 $(x>n)$，那么

$$z_x=\frac{1}{2}z_{x-1}+\frac{1}{4}z_{x-2}+\cdots+\frac{1}{2^{n-1}}z_{x-n+1}$$

关于 z_x 的生成函数是

$$\frac{\psi(t)}{1-\frac{1}{2}t-\frac{1}{4}t^2-\cdots-\frac{1}{2^{n-1}}t^{n-1}}=\frac{\frac{1}{2}\psi(t)(2-t)}{1-t+\frac{t^n}{2^n}}\tag{2-6}$$

，

其中 $\psi(t)=a_0+a_1t+\cdots+a_nt^n$，游戏的条件是 $a_0=a_1=\cdots=a_{n-1}=0$，$a_n=2^{-(n-1)}$。

拉普拉斯注意到在不多于 x 盘后，游戏结束的概率的生成函数等于式 2–6 被 $1-t$ 除，由此得到 $P(x<+\infty)=1$，利用一些相当复杂的有限差分方程，他也考虑了与这个游戏有关的一些额外的问题。与此问题相似的问题由拉普拉斯本人在 1773 年的论文中已考察过，所用的方法仍然是借助有限差分方程的方法。

(8)连续事件的概率。

令一个事件发生的概率是 q，所求的是这个事件在不超过 x 次试验中连续发生 i 次的概率($x>i$)。这个问题的解由棣莫弗给出，当没有给出证明。孔多塞在 1785 年的论文中也曾给出证明。

在这里拉普拉斯依靠有限差分方程给出的解与孔多塞的解法相似。更进一步，在方法论上，它与所阐述的第七个问题的解法相似：首先，拉普拉斯发现了关于在最后 i 次试验中连续事件发生的概率的生成函数 $f(t)$；然后，用 $1-t$ 去除 $f(t)$，并且计算 t^{r+i} 在商中的系数，这样他就得到了连续事件发生不超过 $r+i$ 次实验的概率。一个相似的问题是：在不超过 $x(x\geq n)$ 次投掷中，一个 n 面的骰子的所有面以一个特定的顺序发生的概率，拉普拉斯在早期已用简单的微分方程的方法解决了。

拉普拉斯也解决了两个相关的问题。第一个是赌徒 A 在任何一单局中赢的概率是 q，其决心赢连续 i 局。问这样一个事件正好发生在 x 局后的概率 y_x 是什么?它总归要发生的概率是什么?对于这个问题，拉普拉斯的解法是：记第二个赌徒 B 在任一盘中赢得的概率为 $p=1-q$，并且 B 正好在 x 盘后连续赢得 i 盘的概率为 y_x，拉普拉斯得到一个关于未知数 y_x 和 y_x' 的有限差分方程体系，并且推出相应的生成函数 $u(t)$ 和 $u'(t)$。把这些函数用 $1-t$ 去除，就解决了这个问题的第二部分。

另一个问题是：让每一个参加者如果输掉一盘，就拿出一个硬币，最后赢者拿走所有存放的钱，那么 A 获得的数学期望是什么?拉普拉斯记 A 赢得的数学期望为 $\sum_x xy_x$，则 A 损失的数学期望为

$$P\sum_x[1-\sum_x y_{x-1}-\sum_x y'_{x-1}]$$

，

括号中表示要玩 x 盘的概率。在这里，拉普拉斯的推理是错误的，其中他没有考虑到 A 赢(或输掉)分隔的盘的可能性。

第三章：事件无限次乘积结果的概率的规律。

第三章讨论概率的界限，即事件不定次乘积结果的概率规律。标准问题是两个事件a，b，概率分别为p和$1-p$；则重复试验$x+x'$次中，a出现x次，b出现x'次的概率是二项式$[p+(1-p)]\varepsilon$展开式中的第$(x'+1)$项；而当x，x'都非常大时，拉普拉斯给出了近似计算公式。从公式中指出了两种界限，一是和事件a的先验概率有关，二是和发生事件a的次数与总次数之比有关。当试验次数增大到无穷时，两种界限趋于一致，概率成唯一的。拉普拉斯曾在早年(1781)用男孩和女孩出生数之比值来研究先验概率(见原始文献)。

从本章开始，拉普拉斯开始转向研究试验次数是无限的情况，这样不可避免地涉及极限的理论。在本章的一开头他就阐述了关于极限理论的观点："测量关于事件增加或者减少的方式独立地收敛于极限(始终随着概率的增加而趋向于极限。这个增长的限定和这个极限是机会的分析中尤其值得关注和更应该精细化的地方"。在这里拉普拉斯已认识到概率理论中所用的极限的推理与当时所理解的一般极限的思想有所不同，正如 E.C.Molina 所说"拉普拉斯在他的思想中，对于在纯数学中所用的极限和概率的频率定义基于其上的极限概率有根本的差别。"

在系统地阐述了他的极限观点之后，拉普拉斯开始给出了他的概率理论中最有价值的结论之一的证明，这个结论我们现在称为棣莫弗-拉普拉斯极限定理。在这里拉普拉斯以传统的二项式开始：假设两个事件A和B各自发生的概率是p和$1-p$，在试验$x+x'=n$次中，A发生x次而B发生x'次的概率是二项式$[p+(1-p)]^n$展开式中第$x'+1$项

$$\frac{(x+x')!}{x!x'!}p^x(1-p)^{x'},$$

所以，在n次试验中，A发生$x-l$次而B发生$x'+l$次的概率，即远离最大项为l的项为

$$\frac{n!}{(x-l)!(x'-l)!}p^{x-l}(1-p)^{x'+l},$$

其中$n=x+x'$。

拉普拉斯应用斯特林公式给出了这个公式的估计值

$$\frac{\sqrt{n}e^{-nl^2/(2xx')}}{\sqrt{\pi}\sqrt{2xx'}}\{1+\frac{nzl}{xx'}+\frac{l(x-x')}{2xx'}-\frac{l^3}{6x^2}+\frac{l^3}{6x'^2}\},$$

z是一个处于$x/(n+1)$和$(n-x)/(n+1)$之间的一个量，即是比单位 1 小的一个模，拉普拉斯计算了对称位于这个展开式中间项两边的项的和，它等于

$$\sum\frac{2}{\sqrt{2\pi\sigma}}e^{-l^2/(2\sigma^2)},$$

其中$\sigma^2=\dfrac{xx'}{n}=n\dfrac{x}{n}\dfrac{x'}{n}$。

至此，拉普拉斯的推演过程和结果与棣莫弗的结果和推演仍然完全相同。接下来拉普拉斯开始用 Maclaurin-Euler 求和公式计算出所有这些成对的和，正是在这里，拉普拉斯把他的方法与棣莫弗的推演方法区别开来。这个定理的最后形式是

$$P\{-\frac{1}{n}\le\frac{\mu-np-z}{n}\le\frac{1}{n}\}=\frac{2}{\sqrt{\pi}}\int_0^{\frac{l\sqrt{n}}{\sqrt{2xx'}}}e^{-t^2}dt+\frac{\sqrt{n}}{\sqrt{2\pi xx'}}e^{-\frac{l^2n}{2xx'}},$$

这里μ是在n次试验中事件A发生的次数，且$x=np+z$，$x'=nq-z$。

如果$z=0$，上述公式能够写成下列的形式

$$P\{-c\le\frac{\mu-np}{\sqrt{npq}}\le c\}=\frac{2}{\sqrt{2\pi}}\int_0^c e^{-t^2}dt$$

但是在这里应该强调的是：拉普拉斯引入的是 x 和 x'，而不是 p 和 q。如果 p 被看作是区间[0，1]的任意一个(随机)变量，拉普拉斯隐含地说，它可以由贝叶斯的方法求出。

拉普拉斯首先用数个例子说明他的结果，他说如果男孩和女孩出生的比率是 18：17，那么 14000 个出生的孩子中男孩的数目位于 7363 和 7036 之间的概率是多少?在这里，$p=18/35$，np=7 200，σ=59.1 366，利用连续性的正确性，平均偏差和标准差的比是

(7 363−7 200)/59.1 366=2.764 785，

在这个标准差下的概率积分是 0.9 943 038。

接着，拉普拉斯进一步用摸球的问题来说明上述定理的应用，在这里，他引入了偏微分方程。

第一个摸球问题是：一个罐子包含了大数目(n)的白球和黑球，每从中抽取一个球，就用一个黑球(放入)代替。所求的是在 r 次抽取后，在罐中仍有 x 个白球的概率 y_{xr}。

拉普拉斯求出一个方程

$$y_{xr+1}=\frac{x+1}{n}y_{x+1r}+\frac{n-x}{n}y_{xr},$$

并且假设 $x=nx'$，$r=nr'$，$y_{xr}=y'_{xr}$。

他继续向前推进出一个类线性偏微分方程

$$\frac{1}{n}\frac{\partial y'_{xr'}}{\partial r'}=\frac{x'}{n}\frac{\partial y'_{xr'}}{\partial x'}+\frac{1}{n}y'_{x'r}$$

他求出的解是 $y'=e^{r'}\varphi(x'e^{r'})$。他这样解释这个问题：这两个罐子起先被两种颜色的球充满，这时抽出白球和黑球的概率各为 p 和 q，那么罐子中白球的起先的数量可以根据中心极限定理(相当于计算 y'_{r0})而计算出来。

第二个问题是：两个罐子中各包含 n 个球，有白球和黑球两种颜色，每种颜色球的总数是 n。从每一个罐子中抽出一个球放入另一个罐子中，求在第 r 次这样的操作后，A 罐中包含 x 个白球的概率是什么?

这个问题的提出和最早的解的给出应归功于 D. 伯努利。后来拉格朗日和 Malfatti 也都曾给出各自的解。拉普拉斯本人在他早期的论文中解决过这个问题，在这里他给出的问题与原来的解放本质上是一致的。这个问题之所以值得关注是因为后来人们发现拉普拉斯在这里的工作与后来称为布朗运动的一种随机过程的思想有关。

最后的问题是一个更一般化的问题。假设有 n 个罐子，其中每一个罐子中有大数目的球：白球和黑球两种，起初白球和黑球的比在这些罐子中是不同的。从第一个罐子中抽取一个球放入第二个罐子中，将第二个罐子摇匀，再从第二个罐子中抽出一球放入第三个罐子中，这个过程持续下去，直到从最后一个罐子中抽取一球放入第一个罐子中。然后这个过程重复地一次一次进行下去。通过类似的分析，拉普拉斯发现每一个罐子中的白球数趋向于所有白球的平均数。白球与黑球在每一个罐子中的数目比将以同等而结束。拉普拉斯认为这个问题分析表明了中心极限定理的正确性，并且在一定程度上演示了自然界的隐藏的秩序。

第四章：关于大数次观察的平均结果的误差概率和最好平均结果的概率。

第四章讨论误差的概率问题，是他 30 年来有关误差方面工作的总结。着重讨论了两个问题：一是大量观测资料平均值的误差在一定范围内的概率；二是更有利的平均值误差在一定范围内的概率。在这里已提出最小二乘法的原理，以后由勒让德和高斯最后完成方法体系。

观察的误差的思想由来已久，甚至托勒密就已意识到观察的误差，并建议怎样安排观测。在 11 世纪，A.Biruni 在文章“测地学”中已经注意到天文观测和计算中的随机误差。在对于观测的数学处理中，他用了一个量化方法以反映通常的随机误差的随机性质。但是误差理论被建立起来是在 18 世纪中期，主要由辛普森和兰伯特给出的。在 18 世纪后期，D.伯努利做了一些重要的工作，特别是他把观察到的误差分为随机误差(正态地分布)和系统误差(不变)。不过在拉普拉斯之前不存在下面关键问题的一般解。例如，若处理子午线弧度的测量：给出一些方程

$$a_i x + b_i y + c_i z + \cdots + l_i = 0 (i = 1,2,3,\cdots,n),$$

其中 $m(m<n)$ 个未知的 $x,y,z,\cdots$，现在需要选择合适的解使得残差尽可能地小，并且估计出由于各个自由项 l_i 的误差所引起的这个解的误差。

在 1774 年和 1776 年的两篇文章中，拉普拉斯就尝试利用一个新的数学工具——密度函数研究了各种可能的方式，并比较了各种对于未知参数的估计的精确性不等的标准。其中，有一个条件后来称为了他的主要标准，即绝对期望

$$a_i x + b_i y + c_i z + \cdots + l_i = 0 (i = 1,2,3,\cdots,n),$$

的最小化。在 1810 年和 1811 年的论文中，这是他估计误差的最重要的标准，此外他还提出了最小二乘法。这两篇文章的内容成为第四章的主要内容。《分析概率论》的第四章是拉普拉斯在误差理论中工作的一个高峰和总结。Todhunter 在他的历史著作中对这一章的评论具有代表性，他说：“第四章的内容是拉普拉斯的概率论中最重要的部分，大概也是最难读懂的内容，它包含了极有价值的理论，称之为“最小二乘法”，……拉普拉斯在这一章的过程非常独特，很少有人能够读懂或者理解这一部分的内容，如果不把它们翻译成更通常的语言”。所以为了使人能更好对理解则一部分的内容，很多人致力于将拉普拉斯的理论通俗化和简单化。在 1872 年和 1832 年的《天文年历》上发表了泊松关于观察的平均结果的概率的最有价值的两篇文章，其内容是对《分析概率论》的第四章的一个评论，它们也被收录在 1832 年泊松的“关于判决概率的研究”的第四章中。从此，泊松这部分工作成为阅读《分析概率论》第四章的一个最好的导言，为将来在这个领域的旅行者铺就了一条轻松与安全之路，成为十九世纪大多数概率论书籍中介绍这一部分内容的一个先导。最小二乘法是拉普拉斯概率论的主要内容之一，以它为核心的误差理论在 19 世纪替代古典概率论二成为新的研究核心。所以，有必要介绍一下拉普拉斯在这一章中处理最小二乘法的主要思路。

假设观察赋予一个未知的元素以值 $a_1,a_2,a_3,\cdots$ 等，令 $\phi(x)$ 为一个误差 z 的可能性的函数，这个函数被假设为在每一次观测中是相同的。现在要求这个元素的真值是 x 的概率，所以在每次观察中误差是 $a_1-x,a_2-x,a_3-x,\cdots$。

令 $P=\phi(a_1-x)\phi(a_2-x)\phi(a_3-x)\cdots$，然后根据逆概率的一般法则，真值处于 x 和

$x+dx$ 之间的概率是 $Pdx/\int Pdx$。假设分母上的积分可以在所有 x 可达到的值上积分。

设 H 是满足等式 $H\int Pdx=1$ 的值，令

$$y=H\phi(a_1-x)\phi(a_2-x)\phi(a_3-x)\cdots,$$

拉普拉斯想象以 y 和 x 为坐标轴的曲线，他说我们所取的观察的平均结果的值是使得平均误差为最小的值，每一个误差假设为正的。他展示了这一点正是平分这条曲线的那一点，这是他认为是那个真值要么超过一点要么少一点的最好的平均结果。

然而，拉普拉斯开始解决下面的问题：如果把由观察所提供的结果的平均值作为最有可能的结果，并假设正的误差和负的误差是等可能的，且误差可能性的函数在每次观察中是相同的，那么，所隐含的误差可能性的函数是什么？

设误差 z 的可能性的函数为 $e^{-\psi(z^2)}$，它只涉及到一个假设：正的和负的误差是等可能的，这样上面的 y 的值就成为

$$He^{-\sigma},$$

其中 $\sigma=\psi(x+a_1)^2+\psi(x+a_2)^2+\psi(x+a_3)^2+\cdots$。

为了得到最可能的结果，我们必须求出 x 以使得 σ 最小，这样就得到方程

$$(x+a_1)\psi'(x+a_1)^2+(x+a_2)\psi'(x+a_2)^2+(x+a_3)\psi'(x+a_3)^2+\cdots=0.$$

假设平均值是最可能的结果，设在 s 次观察中有 i 次结果是 a_1，有 $s-i$ 次结果是 a_2，那么上述的方程就变为

$$i(x+a_1)\psi'(x+a_1)^2+(s-i)(x+a_2)\psi'(x+a_2)^2=0,$$

在这种情况下平均值是 $\frac{ia_1+(s-i)a_2}{s}$，用这个值去替换方程中的 x，得

$$\psi'\{\frac{s-i}{s}(a_1-a_2)\}^2=\psi'\{\frac{i}{s}(a_1-a_2)\}^2$$

这个式子并非对所有的 i/s 和 a_1-a_2 的值都成立，除非 $\psi'(z)$ 是独立于 z 的，即 $\psi'(z)=c$，所以 $\psi(z)=cz+c'$，其中 c 和 c' 是常数。

这样，误差 z 的可能性的函数就是形如 Ce^{-z^2}，因为误差必定位于 $-\infty$ 和 $+\infty$ 之间，所以

$$C\int e^{-z^2}dz=1,$$

这样 $C=\frac{\sqrt{c}}{\sqrt{\pi}}$。

在一个变量的情况下，由最小二乘法所给出的结果与所取的平均值相同，因为，当

$$(x+a_1)^2+(x+a_2)^2+\cdots+(x+a_s)^2$$

取最小值时，

$$x=\frac{a_1+a_2+\cdots+a_s}{s}.$$

这样上述推导的结论：由观察所得到的结果的平均值是最可能的结果，与假设“最小二乘法将提供最可能的结果”，这两个命题是等价的。而最小二乘法在拉普拉斯之前已由勒让德和高斯两人提出来。拉普拉斯所指出的这一点具有重要意义，长期以来，人们从实

践中体会到算数平均值的优良性，可并无理论上的根据，以算数平均值作为理论的出发点就显得底气不足，而拉普拉斯把这断裂了的一环连接了起来，从数学上给出了证明，并且拉普拉斯明确赋予这个方法极其重要的地位："不管误差的频率定律是什么，必须说明的是这种方法(最小二乘法)是应当优先选择的"。

在第四章中，拉普拉斯还证明了只要做到适当的中心化和正态化，且随机变量ξ是独立的(拉普拉斯没有提到这个条件)和在一个有下限和上限的区间上是独立的，那么和$\sum\xi_i$，$\sum|\xi_i|$，$\sum\xi_i^2$，$\sum q\xi_i$的极限分布是相同的，当时拉普拉斯没有在这里给出证明，因为这是他以前文章中的某些结果的推论。所以，除了最小二乘法之外，拉普拉斯对于误差理论的贡献的意义还在于他阐明了异常丰富的观察误差是多个独立的基本误差的和一个结果这样一种思想，如果这些误差一致地小，那么在一般的条件下，观察误差的分布必定接近正态的分布，这种思想是对现代的数理统计影响深远的一个观点。

第五章：概率理论在对于自然现象和它们的原因的思考中的应用。

第五章讨论概率计算应用于各种现象及其原因的研究。主要例子是气压在一天内的变化。正常情况下是上午 9 时气压最高，下午 4 时最低；以后上升，到晚上 11 时形成较小的高峰，然后降低到早晨 4 时为止。拉普拉斯通过正常情况的概率计算，指出产生此种周期变化的原因是太阳作用，并定出其平均范围。他试图从数学上解决这个问题，用大气潮汐作为第二个原因，但由于资料缺乏而未完成。他在《天体力学》第二卷第四册中提到这点。他还试图研究心理学现象，希望从大量观测中定出电磁作用在神经系统中的影响。另外还考虑过一事件推断出的原因概率和未来事件概率。这是他在 1774~1786 年间有关原因概率、逆概率用于人口论的再创作。并以那不勒斯、巴黎和伦敦的人口调查数字为基础，估计出法国的人口数及其或然误差。

在本章的开头，拉普拉斯又一次明确地强调了他对于概率这个学科的本质的理解，即保持在他的头脑中的始终是对于概率在物理、天文和自然科学中的一般应用："在这一章里，确定性状态现象的存在性，以及对于它们的唯一的测量是概率演算的目标。"

在这里，拉普拉斯讨论了概率应用于现象本身及其原因的探索，这种探索应建立存在于世界所有的复杂性中的数据的物理意义。这种方法提供了对应于知识和无知、科学和自然的这个学科的相对性。在误差的分析中，一直认为是确定的存在性和范围却成了不确定演算(概率演算)的目标。气压计的日差，经过长期和频繁的观察一般显示出在早晨 9：00 达到最高，在下午 4：00 达到最低，然后在晚上 11：00 升到一个小小的高峰，然后再回落，直到早晨 4：00。拉普拉斯通过概率计算发现这种每日的图像是由于某些规则的原因，是因为太阳运动的影响，拉普拉斯求出了这种变化的平均程度。

对于概率论的可应用性，他充满了信心——概率可以应用医药问题和经济问题中，甚至可以应用到道德问题中去，因为被重复许多次的原因的操作就像在物理学中一样有规则。但是，在这一章中的最后，拉普拉斯没有进一步提出应用的例子，他就用今天在任何一本教科书中都能看到的一个非常著名的问题——蒲丰投针问题而结束了本章。

法国科学家蒲丰(1707~1788)以他的三十六卷关于自然历史的著作而出名。1777 年，他在题为"道德算术"一文中提出了这样一个问题：在一个平面上画有一组间距为a的平行线，将一根长为$l(l<a)$的针任意投掷在这个平面上，问针与任一平行线相交的概率是多

少?蒲丰证明了该针与任一平行线相交的概率 p 是 $\frac{2l}{\pi a}$。拉普拉斯在这里首先指出，通过多次投针实验，并记下针与线相交的次数，从而得到 p 的统计估计值，那么利用上述公式即可求出 π 的近似值(这种计算的精确性是相当低的，1850 年，瑞士的伯尔尼的一位叫沃尔夫的人在投掷 5000 多次后得到 π 的近似值为 3.1596。到目前为止用这种方法得到的最好的 π 的近似值是意大利人拉泽里尼，他在 1901 年投掷了 3 408 次，得到 π 的值精确到 6 位小数)。拉普拉斯还把蒲丰的问题推广到两组互相平行等距离的坐标方格的情况。如果这两组平行线间的距离分别为 a 和 b，那么投掷长度为 l($l < a$ 和 $l < b$)的针与任一直线相交的概率为

$$p=\frac{2l(a+b)-l^2}{\pi ab}.$$

拉普拉斯还提出了一个发现针的最优长度的相关问题，以后这个问题被伯恩斯坦在其 1946 年的《概率论》一书中推广为包括 $l \geq a$ 的更一般的情形，例如相交于两个或更多个线的可能性的问题，以及投掷有一定宽度的物体而不是针等。在这里拉普拉斯并没有费心去证明这个问题，而是认为从以下条件中推出的 l 是最合适的

$$\frac{4l}{\pi a}=1.$$

第六章：关于原因的概率和从过去的经验所导致的未来事件的概率。

第七章：存在于假设相等的几率之间的未知的不等式。

本章专门讨论处于两个几率之间的未知的不等式的影响的估计。这两个几率开始时被假设为在被研究的随机现象中彼此相等。

例如，令在投掷一个硬币时正面发生的概率等于 $p=\frac{1+\alpha}{2}$，其中 $\alpha \neq 0$ 未知，那么在 n 次投掷中出现 n 个正面的概率是什么?显然是 $\frac{(1+\alpha)^n}{2^n}$。

第八章：寿命、婚姻和一般的结合的平均持续时间。

本章专门讨论人口统计学，是第六章所讨论问题的继续和补充。

第一个问题，生命的寿命。不管在方法论上还是在数学中，平均寿命的真实值的计算和它的经验的可能的误差的计算都不同于第四章中对应的计算(第四章是专门讨论误差理论的)，第四章是专门讨论误差理论的。

第二个问题是婚姻的平均持续时间问题。

第九章：基于将来事件发生的概率的收益问题。

第十章：道德期望。

第八、九、十章讨论了大量社会现象的概率问题，例如平均年龄、年金、保险、婚姻期限、各社会团体存在期限等。第十章主要讨论了与数学期望(平均值)不同的可能期望值的计算，比原来的伯努利公式更有用。

第十一章：证据的概率。

典型例子是抽签问题。设在袋中有带号码的签条，抽出一根给见证人看后，他说号码为 n。这是真的吗？拉普拉斯用逆概率来讨论这个问题，估计见证人可靠性的概率。共有四种可能性：

(1)见证人既未说谎也未看错。

(2)他没有说谎但看错了。

(3)他说谎但没有看错。

(4)他又说谎又看错。

拉普拉斯还推广到几个见证人，情况更复杂。最后结论是，对相信者保证得越多的见证人，其可靠性概率越小。他后来在《概率的哲学导论》中，举了数字例子；并揭示了帕斯卡论证上帝存在性的错误。

书末有三个补充，都是书中一些公式的具体证明或推广。另外有后来增加的四个附录。

1816 年完成的附录一的标题是“关于概率在自然哲学中的应用”。实际上主要讨论社会问题，以一判罪为例。拉普拉斯认为犯罪的原因概率中，已知或已假定其先验概率；并认为陪审员或法官的可靠性概率在 1/2 到 1 之间。他计算出陪审团中几人投赞成有罪票得到的判决差错的概率。如陪审团为 8 人，5 人投有罪票，差错概率为的 65/256，由此讨论了陪审团组成方案。

1817~1819 年间，拉普拉斯把概率论用于提高大地测量资料的精度，相应研究结果归纳为附录二、三，加在《概率分析理论》第三版中。这两个附录在统计学历史发展中有重要意义，实质上就是拉普拉斯的误差理论，用最小二乘原理使仪器误差和观测误差极小化。附录二的标题为“关于概率计算在测地活动中的应用”。主要提出自己的统计方法，并同其他方法进行比较。附录三的标题为“概率测地公式在巴黎子午线测量中的应用”。他的助手 J.B.J.德朗布尔(Delambre)等在 1796 年整理时只用了 27 个三角网资料，而拉普拉斯用了全部 700 个三角网原始资料。算出了不同子午线长度量级的误差概率，得出巴黎子午线长度的最优值。

1825 年写完的附录四是关于生成函数的四点小补充。拉普拉斯于 1825 年撰写完毕。

4.关于《分析概率论》的评论

从以上对《分析概率论》的分析来看，毫无疑问这部著作应当被看做当时在这个领域里达到的顶峰之作，因为拉普拉斯综合汇集整理了当时几乎所有已知的概率和统计的问题，而且他自己所有的关于概率理论的工作和观点都汇集在这部著作中，可以说，这是他的关于概率论知识一生工作的总结。所以，这部著作足以代表和体现拉普拉斯的概率理论的特点和成就。至此我们可以对拉普拉斯的概率工作的几个方面作一个评价。

首先，在拉普拉斯思想的深处，概率论与其说是一门纯数学，不如说是一门有用的自然科学更合适。对于数学概率的这种理解为拉普拉斯的概率论打上了两个鲜明的特征：一个是研究的对象集中于概率论的应用领域；另一个是把应用的成功看作检验它的正确性的标准。所以在拉普拉斯的概率论工作中，对概率在各个领域中的应用研究成为他的主要目标：这些领域包括在机会性游戏中的应用；在保险、天文、医药、证言的可靠性、法庭判决的精确性、价值的经济理论，从已知的现象推测未知的原因等，这一系列的问题实际上都是建立在“概率论是应用科学”这一信念基础上的。拉普拉斯像是一个勇敢的探险者，他急于开拓概率论所延伸的各种应用领域，他迫不及待地在每一个壕沟上架起通向未知领域的桥梁，他时常为自己的一些成功所鼓舞，但是无暇暂时停下来检查一下这些桥梁的基础是否牢固。泊松曾经对作为自然科学家的拉普拉斯与作为纯数学家的拉格朗日作过恰当的比较：

“这是广阔的问题范围，其中占据了这些伟大的领袖人物的精力，对于这些问题的思考他们是从完全不同的角度出发的，并且仍然没有穷尽课题。在这两个天才中的确存在着一个区别，任何研究他们作品的人都会意识到这一点。不管是关于月天平动的问题，还是数论中的问题，拉格朗日在他处理的问题中似乎更经常地看到它们所化归的数学，然后花很高的代价使他的公式达到优雅，使他的方法达到一般性。

而对于拉普拉斯正相反，数学的分析在他看来是有着最广阔的应用的一个工具，他经常通过使特殊的方法以从属于特殊问题的需要。大概后代们将这样来判断：一个是伟大的数学家，而另一个是通过把最高贵的数学引进来以服务于理解自然的科学家。就这样，拉普拉斯给我们提供了毛心管作用的理论，对此他求出了应用于大数次观察的不同过程的计算的概率的度量；他表述了涨潮和落潮的公式；……”

拉普拉斯以应用为中心的特点在《分析概率论》的每个方面体现出来：在体系的安排上，拉普拉斯并没有试图给概率论一个像欧几里得的《几何原本》那样一个以逻辑顺序安排内容，《分析概率论》并没有任何明晰的逻辑演绎结构，更像一般意义上的一篇篇独立的论文的汇总。虽然他所用的数学解题方法之多令人目不暇接，他称得上是一个聪明的解题高手，然而他并没试图探求一般的思想和方法。综观此书的内容，可以明显看到它是以这个学科发展进程的历史，尤其是按照概率论在某个领域中的应用的顺序来安排的。在内容上他也没有将数理统计从概率论中分离出来。他的概率论著作称得上是集概率、统计、物理等当时几乎与所有知识领域有关的内容的一个汇集。在写作风格上，他更不会对概率论这一学科进行精雕细刻。尽管概率论在拉普拉斯一生的研究中占据了重要的分量，但是概率论对他的吸引力并不是其自身内容和体系的优美，而是概率论自身巨大的应用价值。所以拉普拉斯不会为概率论学科的系统体系的完善而付出努力。在《分析概率论》的很多地方，他并没有花时间去解释他的分析步骤和雕琢他的作品，如果他知道或感觉结论是正确的，他常以“显而易见”就避免了完整的讨论，或者仓促得出一些不甚严格甚至错误的结论。

其次，拉普拉斯概率研究中的另一个特点是他的概率研究大大受益于对数学历史的透彻了解。他说的一段话就包含了一种对历史精神的精辟理解：“我最大的乐趣是研究发明者的进展情况，看看他们的天才怎样对付他们遇到的障碍，和怎样克服这些障碍……”。的确，我们看到拉普拉斯的概率理论从早期就受惠于雅各布·伯努利、棣莫弗、拉格朗日、达朗贝尔和孔多塞等人的数学工作，他研究的每一个概率问题都可以追索到他的同时代人或更早的前辈那里去，但是，一如既往，拉普拉斯在关键的地方很少提到这些问题的渊源，这种风格在他的《分析概率论》中体现的更明显。这种状况导致了后人常常对拉普拉斯在概率论的历史上究竟是一个创造者还是一个综合者的争论。

从上面的分析中可以得出这样一个结论：拉普拉斯在(他的意义上的)概率论方面既是一个伟大的创新者，又是一个全面的综合者。在创新方面，他引进了生成函数的理论以及特征函数的理论的基本思想，并将之作为一个基本工具用在中心极限定理的内容中。他发展了现在所谓的贝叶斯定理以作为统计推断的基础，就像争论的那样，不管拉普拉斯是否独立地发现了贝叶斯定理，正是拉普拉斯抓住了这个定理，并将它发展成为远远超出贝叶斯本人所想象的科学探索的一个有力工具；他对误差理论和最小二乘法做出了巨大的贡献，这里他的主要框架是关于具有对称分布的独立的和等分布的残差的一个线性回归的基本结构。他利用了渐近发来判断最小二乘法的应用，这一点不同于高斯于 1909 年首先用一

个固定的样本容量和正态地分布的残差。两人都用了现在所称的决策论中的方法。在人口统计方面他推广了丹尼尔•伯努利早期对于疾病和死亡率方面的影响，以及为了预防传染病而实行的免疫接种方面的问题，他在人口论中依据性别出生的比率而发展了置信区间的思想。他还改进了诸如“点数问题”、“赌徒输光问题”等古典概率论中早期的许多问题。他严格地拓展了雅各布·伯努利弱大数定理和棣莫弗的中心极限定理结果到具有有限子集的离散的随机变量的情形和非严格的等分布的随机变量的一般情景。并且，后继于孔多塞，他在法庭判决的制定的概率分析方面做了重要的工作，尽管这一部分内容是以后受抨击最多的地方，但是不可否认，以后的犯罪统计学的产生于拉普拉斯在这方面的大力鼓吹和提倡有着不可分割的渊源关系。

由此看到，现代概率论和数理统计的许多重要分支的思想大都可以在《分析概率论》中找到它们的原始萌芽。这部著作不仅仅是一个数学界的元老对一个学科的经典领域的所有成果一生工作的汇总，而更重要的是对一些陈旧问题的新的创造，并试图将它们带入原来只是想象的但没有具体实施的领域中。所以，用“近代概率论理论之父”的称号来描述拉普拉斯在概率论历史中的地位并不过分。

第三，拉普拉斯的概率论内容是他那个时代的数学特征的典型代表：哲学上的决定论思想与数学工作相混合、18 世纪的分析特征与深奥微妙的科学认识论相结合、个人出人头地的雄心与精湛细致的数学技巧的展示相结合。特别是一直贯穿在拉普拉斯的概率论研究中的一个理想更是代表了启蒙运动时期的大多数知识分子的主要意图和理想：将社会科学也像自然科学一样建立在公正严密的数学基础之上，正像他又一次在《概率的哲学导论》所说的：真理、正义、人性——有一个唯一建立在人和他的同类以及和自然的关系基础之上的社会秩序的永恒定律。这种秩序的保持对于人类社会的必要性就像万有引力定律在自然的秩序中存在一样。由于人类的无知，概率论是接近这个目标的一个主要手段。所以，拉普拉斯概率理论是一条漫长而曲折的思想潮流的结果，是西方文化的产物。它的发展与十九世纪欧洲的社会政治、经济、科学、宗教信仰等状况息息相关。

最后，拉普拉斯因为站在了当时概率论发展的顶峰而得以对这门学科的前景高瞻远瞩。在他那个时代，可能他关于概率对于世界的意义的理解胜过任何其他人，用他自己的话说就是：

“对于生活中的大部分，最重要的问题实际只是概率问题。严格来说，你可以说差不多我们所有的知识都是不确定的，只有一小部分我们能确定地知道。甚至数学科学本身，归纳法、类推法和发现真理的首要手段都是建立在概率论的基础上的。因此，整个的人类知识系统是与这一理论相联系的……很显然，起源于机会游戏思考的概率，应该成为人类知识中最重要的一门学科。”

只要浏览一下今天的报纸，看一看今天的电视，听一听广播，我们会看到在某种程度上概率和统计的语言的确已经成为人类生活中重要的一部分。不仅仅在自然科学中，概率论在社会科学的涉入也在增加，这一切不妨看作是拉普拉斯的预言的部分实现。

2.3 高斯和泊松的概率论工作

高斯（Carl Friedrich Gauss，1777~1855）是德国数学家、物理学家、天文学家。1777

年 4 月 30 日生于不伦瑞克，1855 年 2 月 23 日卒于哥廷根。

高斯的祖父是农民，父亲是园丁兼泥瓦匠。高斯幼年就显露出数学方面的非凡才华：他 10 岁时，发现了

$$1+2+3+4+\cdots+97+98+99+100$$

的一个巧妙的求和方法；11 岁时，发现了二项式定理。高斯的才华受到了布伦瑞克公爵卡尔·威廉（Karl Wilhelm）的赏识，亲自承担起对他的培养教育，先把他送到布伦瑞克的卡罗林学院学习（1792~1795 年），嗣后又推荐他去哥廷根大学深造（1795~1798 年）。

高斯在卡罗林学院认真研读了牛顿、欧拉、拉格朗日的著作。在这时期他发现了素数定理（但未能给出证明）；发现了数据拟合中最为有用的最小二乘法；提出了概率论中的正态分布公式并用高斯曲线形象地予以说明。进入哥廷根大学第二年，他证明了正 17 边形能用尺规作图，这是自欧几里得以来两千年悬而未决的问题，这一成功促使他毅然献身数学。高斯 22 岁获黑尔姆斯泰特大学博士学位，30 岁被聘为哥廷根大学数学和天文教授，并担任该校天文台的台长。

高斯在数学世界"处处留芳"：他对数论、复变函数、椭圆函数、超几何级数、统计数学等各个领域都有卓越的贡献。他是第一个成功地运用复数和复平面几何的数学家。他的《算术探究》一书奠定了近代数论的基础；他的《一般曲面论》是近代微分几何的开端；他是第一个领悟到存在非欧几何的数学家；是现代数学分析学的一位大师，1812 年发表的论文《无穷极数的一般研究》，引入了高斯级数的概念，对级数的收敛性作了第一次系统的研究，从而开创了关于级数收敛性研究的新时代，这项工作开辟了通往 19 世纪中叶分析学的严密化道路。在数学中以他的姓名命名的有：高斯公式、高斯曲率、高斯分布、高斯方程、高斯曲线、高斯平面、高斯记号、高斯概率、高斯变换、高斯分解、高斯和、高斯素数、高斯级数、高斯系数、高斯准则、高斯原理、高斯消元法、高斯过程、高斯映射、高斯测度、高斯二次型、高斯多项式、高斯不等式、高斯随机过程、高斯随机变量……拉普拉斯认为："高斯是世界上最伟大的数学家"。

高斯在他一生中，只对一种人感到反感和蔑视：这就是明知自己错了又不承认错误的、佯装有学问的人。他的国家的人民为了缅怀、纪念高斯，特将他的故乡改名为高斯堡，并在他的母校哥廷根大学建立了一座以正 17 边形棱柱为底座的高斯雕像。

1809 年，高斯发表了其数学和天体力学的名著《绕日天体运动的理论》。在此书末尾，他写了一节有关"数据结合"（data combination）的问题，实际涉及的就是这个误差分布的确定问题。

设真值为θ，n个独立测量值为$X_1,X_2,\cdots,X_n$。高斯把后者的概率取为

$$L(\theta)=L(\theta;\ X_1,X_2,\cdots,X_n)=f(X_1;\theta)\cdots f(X_n;\theta), \quad (2-7)$$

其中f为待定的误差密度函数。到此为止他的作法与拉普拉斯相同。但在往下进行时，他提出了两个创新的想法。

一是他不采取贝叶斯的推理方式，而径直把使上式达到最大的

$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$作为$\theta$的估计，即使

$$L(\hat{\theta})=\max_{\theta}L(\theta) \quad (2-8)$$

成立的$\hat{\theta}$。现在我们把$L(\theta)$称为样本$X_1,X_2,\cdots,X_n$的似然函数，而把满足式 2-8 的$\hat{\theta}$称

为θ的极大似然估计。这个称呼是追随费歇尔，因为他在 1912 年发表的一篇文章中，明确提到以上概念并非针对一般参数的情形。

如果拉普拉斯采用了高斯这个想法，那他会得出θ的估计是$X_1,X_2,\cdots,X_n$的中位数$med(X_1,X_2,\cdots,X_n)$，即$X_1,X_2,\cdots,X_n$按大小排列居于正中的那一个（n为奇数时），或居于正中那两个的算术平均（n为偶数时）。这个解不仅计算容易，且在实际意义上，有时比算术平均X更合理。不过，即使这样，拉普拉斯的误差分布大概也不可能取得高斯正态误差那样的地位。原因是X是线性函数，在正态总体下有完善的小样本理论，而$med(X_1,X_2,\cdots,X_n)$要用于推断就难于处理。另外，这里所谈的是一个特定的问题——随机测量误差该有如何的分布。测量误差是由诸多因素形成，每种因素影响都不大。按中心极限定理，其分布近似于正态是势所必然。其实，早在 1780 年左右，拉普拉斯就推广了棣莫弗的结果，得到了中心极限定理的比较一般的形式。可惜的是，他未能把这一成果用到确定误差分布的问题上来。

高斯的第二点创新的想法是：他把问题倒过来，先承认算术平均X是应取的估计，然后去找误差密度函数f以迎合这一点，即找这样的f，使由式 2—8 决定的$\hat{\theta}$就是$\overline{X}$。高斯证明这只有在

$$f(x)=\frac{1}{\sqrt{2\pi}h}e^{-\frac{x^2}{2h^2}} \tag{2-9}$$

才能成立，这里$h>0$是常数，这就是正态分布$N(0,h)$。

使用这个误差分布，就容易对最小二乘法给出一种解释。高斯这项工作对后世的影响极大，它使正态分布同时有了“高斯分布”的名称。后世之所以将最小二乘法的发明权归之于他，也是出于这一工作。我们要从最复杂原理配合不同的约束条件，从理论上推导出概率论中常见的若干概率分布函数。这为分散的概率分布函数找到了统一的，系统的物理的说明，也使概率分布的理论成因的物理背景清楚了一大步。很多常用的概率分布函数都可以从最复杂原理导出，这为理解这些概率分布的成因提供了统一的物理思路。把这个认识吸收到概率论和统计学中会使这些 学科的科学水平提高一步最复杂原理就是有随机性的客观事物（广义几何）都自动使自己内部状态的复杂程度在限制条件下达到最大值。由此引申出另外一个结论：最容易出现的事物是复杂性最高的事物。这既是最复杂原理的由来，也是它正确性的依据。

泊松(Poisson，1781~1840)是法国数学家、物理学家和力学家。1781 年 6 月 21 日生于皮蒂维耶，1840 年 4 月 25 日卒于巴黎附近的索镇。

泊松的父亲是退役军人，退役后在村里作小职员，法国革命爆发时任村长。泊松最初奉父命学医，但他对医学并无兴趣，不久便转向数学。于 1798 年进入巴黎综合工科学校，成为拉格朗日、拉普拉斯的得意门生。在毕业时由于其学业优异，又得到拉普拉斯的大力推荐，故留校任辅导教师，1802 年任巴黎理学院教授。1812 年当选为法国科学院院士。1816 年应聘为索邦大学教授。1826 年被选为圣彼得堡科学院名誉院士。1837 年被封为男爵。著名数学家阿贝尔说：“泊松知道怎样做到举止非常高贵”。

泊松是法国第一流的分析学家。年仅 18 岁就发表了一篇关于有限差分的论文，受到了勒让德的好评。他一生成果累累，发表论文 300 多篇，对数学和物理学都作出了杰出贡献。

在数学方面：美国数学史家克莱茵（Kline）指出："泊松是第一个沿着复平面上的路径实行积分的人"。在他 1817 年的出版物中对序列收敛的条件就有了正确的概念，现在一般把这个条件归功于柯西。泊松对发散级数作了深入的探讨，并奠定了"发散级数求积"的理论基础，引进了一种今天看来就是可和性的概念。把任意函数表为三角级数和球函数时，他广泛地使用了发散级数，用发散级数解出过微分方程，并导出了用发散级数作计算怎样会导致错误的例子。他还把许多含有参数的积分化为含参数的幂级数。他关于定积分的一系列论文以及在傅里叶级方面取得的成果，为后来的狄利克雷和黎曼的研究铺平了道路。

泊松也是 19 世纪概率统计领域里的卓越人物。他改进了概率论的运用方法，特别是用于统计方面的方法，建立了描述随机现象的一种概率分布——泊松分布。他推广了"大数定律"，并导出了在概率论与数理方程中有重要应用的泊松积分 $\int_{0}^{+\infty} e^{-x^2} dx$。他是从法庭审判问题出发研究概率论的，1837 年出版了他的专著《关于刑事案件和民事案件审判概率的研究》。

泊松就三个变数的二次型建立起特征值理论；并给出新颖的消元法；研究过曲面的曲率问题和积分方程。

在数学物理方面：泊松解决了许多热传导方面的问题，他使用了按三角级数、勒让德多项式、拉普拉斯曲面调和函数的展开式，关于热传导的许多成果都包含在其专著《热的数学理论》之中。他解决了许多静电学和静磁学的问题；奠定了偏向理论的基础；研究了膛外弹道学和水力学的问题；提出了弹性理论方程的一般积分法，引入了泊松常数。他还用变分法解决过弹性理论的问题。

在引力学中，他发表了《关于球体引力》和《关于引力理论方程》的论文，引入了著名的泊松方程。他的名著《力学教程》（2 卷），发展了拉格朗日和拉普拉斯的思想，成为广泛使用的标准教科书，在天体力学方面，他研究了关于月球和行星理论以及太阳系稳定性的某些问题，计算出由球体和椭球体引起的万有引力。他 1831 年还发表了《毛细管作用新论》。

泊松一生对摆的研究极感兴趣，他的科学生涯就是从研究微分方程及其在摆的运动和声学理论中的应用开始的。直到晚年，他仍用大部分时间和精力从事摆的研究。他为什么对摆如此着迷？有一个传说，泊松小时候由于身体孱弱，他的母亲曾把他托给一个保姆照料，保姆一离开他时，就把泊松放在一个摇篮式的布袋里，并将布袋挂在棚顶的钉子上，吊着他摆来摆去。这个保姆认为，这样不但可以使孩子身上不被弄脏，而且还有益于孩子的健康。泊松后来风趣地说：吊着我摆来摆去不但是我孩提时的体育锻炼，并且使我在孩提时就熟悉了摆。

在数学中以他的姓名命名的有：泊松定理、泊松公式、泊松方程、泊松分布、泊松过程、泊松积分、泊松级数、泊松变换、泊松代数、泊松比、泊松流、泊松核、泊松括号、泊松稳定性、泊松积分表示、泊松求和法……

泊松是无论在兴趣还是观点方面都是拉普拉斯主要的继承者，他是从多科工艺学校毕业的第一代数学家，是拉普拉斯的学生。泊松是一位精力充沛、研究领域广泛的科学家，他的主要研究领域是数学物理，他曾写了一些关于力学和热学理论的论文，他在这些领域的工作和拉普拉斯、拉格朗日和傅立叶等人的工作方式不一样，他在所有这些领域中作用

只是一个称职的他人工作的拓展者，并非是一个勇敢的新领域的开拓者，在对概率论及其应用的工作中也是如此。

泊松对于概率论的兴趣是受到拉普拉斯关于这个科目的工作的激发。泊松的关于概率论的著作和文章数量不多，其中最有影响也是最有代表性的著作是 1837 年出版的《关于犯罪和民事判决的概率的研究》。这本书的结构和内容上都明显地印有拉普拉斯《分析概率论》的风格和观点的痕迹。此书的前四章主要是在数学和哲学的框架中来阐述，其目的是为以后所要讨论的问题提供一个合理的哲学基础和有效的数学方法。在文章的后面，他才处理与题目有关的问题，这些问题无疑是对《分析概率论》的第十一章中所讨论问题的继续。泊松在这方面的工作主要涉及凯特勒关于法院判决的形成的模型的应用，以及有关 1825 年至 1830 年的关于法国犯罪方面的数据分析。泊松在概率对于法院观念形成的应用并没有新的创造——他的目标也是对孔多塞和拉普拉斯在这方面工作的细化和提供一些更好的解释例子。他和凯特勒一样也注意到某些数据一年一年令人惊奇的稳定性（泊松曾与凯特勒通信，人们猜测他们彼此交换了法国和布鲁塞尔的统计数据）：自杀、偷盗的总和与被指控人员之间的比例令人吃惊地几乎没有什么变化。泊松视被告有罪无罪为作出判决的“未知原因”，有罪无罪的比例就如同男婴与女婴出生数的比例一样稳定，他主要用来自拉普拉斯对于人口统计中原因概率的方法去解释这些统计的规律，就像男婴出生数占稳定的优势的情形一样，这一切都显示了一个恒定原因的操纵。这一点与凯特勒关于“不可改变的”社会规律最终被个体意志和环境的“干扰的力量”所掩饰的观点遥相呼应。

除了以拉普拉斯的概率统计研究作为一个雄厚的基础外，此时法国的社会环境也为他提供了良好的条件。这时从 1821 年到 1828 年作为司法大臣的 Comte de Peyronnot 发起了一场名为“法国犯罪审判的综合审计”的活动，目的是为评价立法和判定机关程序的功效提供一种方法，并作为可以提供“通过决定趋向增加和减少犯罪数目的环境”方案的一种诊断工具。Peyronnot 阐述了他的信念：为了“用实际确定的知识”来代替含糊的理论，精确的统计是现代政府最迫切需要的东西。原始的报告以犯罪的性别和数目、被控有罪的比率、被宣判的各种类型的惩罚的频率等分门别类地统计——所有信息都以表格的形式排列。后来的版本通过考虑到被控人和被定罪人两者的年龄和性别、不能履行责任的陪审员的数目、罪犯的动机、随季节变化的犯罪率等因素逐渐地扩展表格和多种问题的数目，从 1825 年起在法国开始讨论十二个陪审员中有七人赞同的大多数通过一个犯罪判决的法律条款。泊松也积极地参与到这个讨论中，他的目的和拉普拉斯一样是对陪审团的人数多少的影响以及做出正确判决的决定的概率进行分析。

泊松首先解决两个问题：确定任意给定的陪审员将做出正确决定的概率和先于审判确定被告有罪的概率。为了解决这些问题，泊松借助于从 1825~1827 年和 1827~1830 年由 7:5 的多数所宣布的有罪的数目进行对比分析。泊松假设被告有罪的几率是 k ，有罪和无罪是两个原因，各自的概率是 k 和 $1-k$ ；而 k 在大数目的人口中被看作所有的犯罪数与总人口的比率。

泊松假设所有的陪审员都独立地做出决定，并且每个人得到正确决定的概率是 u ，那么，如果下式

$$B(i,n,u)=\sum_{l=0}^{i}C_n^l u^l(1-u)^{n-l}$$

，

被用来表示n个陪审员中不超过i个得出正确判决的概率，那么对于一个有罪的被告定罪的概率是$c_1 = B(5,12,\cdots,1-u)$，对于一个无辜的被告定罪的概率是$c_1 = B(5,12,\cdots,u)$，所以对一个随机选择的一个杯盖来讲，他被判罪的总的概率是

$$\gamma = k \cdot B(5,12,\cdots,1-u) + (1-k) \cdot B(5,12,\cdots,u)$$

泊松关于这个明显的发现的相关部分是：给定大数次μ犯罪实验，其中被告中m个被控有罪的概率γ由m/μ给出估计值。

进一步，如果

$$p = 1 - \frac{2}{\sqrt{\pi}} \int_{\alpha}^{+\infty} e^{-x^2} dx,$$

其中α是正的但相对于$\sqrt{\pi}$很小，p给出了”相对应的未知的确定极限的概率”

$$\frac{m}{n} \pm \alpha \sqrt{\frac{2m(\mu - m)}{\mu^3}} = P(|Z| \le \sqrt{2}\pi),$$

其中Z服从标准正态分布。

这些分析的方法从报纸上与拉普拉斯在《分析概率论》中所讨论的方法相似。不过泊松进一步将这种方法应用到具体的数字分析中去。

泊松在他的这本概率论代表作还给出了与他的名字直接相连的概率论中的两个重要成果。

第一个是现在通常所称的“泊松分布”的概率分布

$$P\ (X = k) = \frac{\lambda^k}{k!} e^{-\lambda} (k = 0,1,2,\cdots)$$

在这本书中，泊松以大数次试验的二项分布的累积形式导出了这个公式，当成功的概率很小时可作为二项分布的一个极限，尽管泊松对此结果没有特别的强调，但这个简略的结果还是引起 Cournot 的注意，他于 1843 年又重新对之叙述，以后才逐渐引起后人的注意。

另一个在现代文献中经常见到的与泊松名字相连的最著名的结果是大数定律。泊松将伯努利大数定律加以推广：

令m是在μ次试验中事件E发生的次数，并假设E在第i次试验中发生的概率是p_i，那么m/μ与$p' = (p_1 + p_2 + \cdots + p_\mu)/\mu$的差会随着$\mu$的增大而趋向于零，即$m/\mu - p'$的绝对值大于任何给定的正数$\varepsilon$的概率将随着$\mu$的增大而趋向于零，也就是说

$$\lim_{\mu \to +\infty} P(|\frac{m}{\mu} - p'| > \varepsilon) = 0$$

在这里泊松对伯努利大数定律的推广是将原来的“等可能性”（即所有的p_i都相等）扩展到不相等的情况，这一点与泊松讨论的各种可能的犯罪率和随后统计系列的稳定性都有关系。

2.4 圣彼得堡数学学派的概率论工作

切比雪夫(Chebyshev,1821~1894)是俄国数学家、力学家。1821 年 5 月 26 日生于奥卡多沃，1894 年 12 月 8 日卒于圣彼得堡。

切比雪夫的左脚生来有残疾，因而童年时代经常独坐家中，养成了在孤寂中看书和思

索的习惯，并对数学产生了强烈的兴趣，特别对欧几里得的《几何原本》中关于没有最大素数的证明所深深吸引，1837 年考入莫斯科大学物理数学系学习，在大学四年级时，他以一篇题为《方程根的计算》的论文，获系里颁发的银质奖章。大学毕业后，他留在莫斯科大学当助教并同时攻读硕士学位，1846 年以题为《试论概率论的基础分析》的论文获硕士学位。其后他到圣彼得堡大学任教。1849 年他以题为《论同余式》的论文获得圣彼得堡大学博士学位，并获圣彼得堡科学院的最高数学荣誉奖。切比雪夫于 1850 年在圣彼得堡大学晋升为副教授，1860 年晋升为教授，1859 年当选为圣彼得堡科学院院士。他还先后当选为法兰西科学院，柏林皇家科学院、意大利皇家科学院、瑞典皇家科学院的外籍院士和伦敦皇家学会会员。1872 年圣彼得堡大学授予他功勋教授称号，1890 年他荣获了法国荣誉团勋章。

切比雪夫在数学的很多方面及其邻近的学科都作出了重要贡献。切比雪夫对数学作出了大量的贡献，在数学中以他的姓氏命名的有：切比雪夫集、切比雪夫交错、切比雪夫点、切比雪夫结点、切比雪夫网、切比雪夫常数、切比雪夫向量、切比雪夫中心、切比雪夫子空间、切比雪夫半径、切比雪夫逼近、切比雪夫函数、切比雪夫方程、切比雪夫系、切比雪夫准则、切比雪夫法、切比雪夫迭代法、切比雪夫参数迭代法、切比雪夫半迭代法、切比雪夫多项式、切比雪夫不等式、切比雪夫定理等等。而其中以他的姓氏命名的定理、方程、多项式、不等式……有多种。

切比雪夫不但研究成果辉煌，而且教学成就卓著，他在圣彼得堡大学执教 35 年间，先后主讲过数论、高等代数、积分运算、椭圆函数、有限差分、概率论、分析力学、傅立叶级数、函数逼近论、工程机械学等十余门课程，他在教学工作中能将自己的精练见解与研究成果融汇于讲课之中，因而深受学生的欢迎。例如，他的学生，著名数学家李雅普诺夫评论道："切比雪夫的课程是精练的，他不注意知识的数量，而是热衷于向学生们阐明一些最重要的观念。他的讲课是生动的、富有吸引力的，总是充满了对问题和科学方法之重要意义的奇妙评论。"由于切比雪夫在圣彼得堡大学几十年来的言传身教，孕育、培养、造就了不少杰出数学家，例如马尔可夫、李雅普诺夫、格拉韦等，从而逐步形成了以切比雪夫为代表的圣彼得堡数学学派。这个学派的特点是：重视基础理论，善于以经典课题为突破口；理论联系实际；擅长运用初等工具建立高深的结果；以大学为基地，科研、教学相结合。

切比雪夫终身未娶，把一生献给了科学教育事业。他去世后，先后出版了他的论文集（1899~1907）、全集（1944~1951）和选集（1955）。1944 年，苏联科学院设立了切比雪夫奖金。

概率论是一门研究随机现象数量规律的科学。随机现象在自然界和人类生活中无处不在，因而大多数的应用研究，无论属于自然科学、工程事务、生产组织、医学或经济，其本质都处于现实过程中随机作用的影响。这个观点强有力地推动了概率论和数理统计的飞速发展。概率论极限理论的创立使其更是如虎添翼，以致发展成今天如此枝叶茂盛、硕果累累的参天大树。概率论的圣彼得堡学派在其中的功绩是不可磨灭的。

俄国数学起步是比较晚的。当古代埃及、巴比伦、古希腊、中国、印度在初等数学领域中取得辉煌成就时，俄国几乎没有什么值得称颂的数学成就。当笛卡尔和费马的解析几何，牛顿与莱布尼兹的微积分称雄欧洲时，俄国数学更显得落后了。此时，俄罗斯没有自己的数学家，没有大学，甚至没有一部像样的初等数学教科书。虽在 1725 年建立了圣彼得堡科学院，但早期的数学院士都是外国人。直至 19 世纪上半叶，俄国才出现了像罗巴

切夫斯基、布尼亚科夫斯基和奥斯特罗格拉茨基这样优秀的数学家。然而在当时他们的研究成果还不足以引起西欧数学家的充分重视。切比雪夫就是在这样的数学文化背景下凸显出的。切比雪夫在概率论、解析数论和函数逼近论领域的突出成就从根本上改变了俄国数学落后的状况，更重要的是他以自己卓越才能和独特风格吸引和培养了一批年轻的俄国数学家，开创了俄国数学的新局面，并形成了一个新的数学学派——圣彼得堡数学学派。

1.俄国概率论的先驱

19 世纪中叶之前，俄国还未将概率论列人数学—物理学专业的教学计划。第一个将概率论应用于保险业和人口统计的俄国人是哈尔科夫大学的教授巴甫洛夫斯基。巴甫洛夫斯基是奥斯特罗格拉茨基的老师，他的概率思想深深地影响了奥斯特罗格拉茨基。1822 年 5 月，奥斯特罗格拉茨基到巴黎学习。他结识了法国许多数学名流如拉普拉斯、傅里叶、泊松、柯西、勒让德等，并与布尼亚科夫斯基在巴黎见面且成为挚友。1828 年春，奥斯特罗格拉茨基回国在圣彼得堡大学开始任教。布尼亚科夫斯基 5 岁丧父，由父亲的战友抚养成人。1820 年到国外深造。他师从于德奥利。德奥利的研究方向是数论，写有不少数理方面的专著，1808 年始为圣彼得堡科学院的通讯成员，并写信举荐布尼亚科夫斯基，而使布尼亚科夫斯基成功当选为圣彼得堡科学院的助手。之后，布尼亚科夫斯基又到巴黎学习两年，结识了一些数学名流。1826 年回到圣彼得堡，1827 年在海军研究院教学，1846 年始在圣彼得堡大学任教。

在巴黎期间，奥斯特罗格拉茨基和布尼亚科夫斯基都与拉普拉斯进行了密切接触，但二人在当时对概率论都没有表示出很大的兴趣。直至在圣彼得堡大学任教后，他们才意识到概率论这一学科的重要性及其在社会问题上的广泛应用。尽管当时泊松之《关于判断的概率之研究》已于 1837 年出版，但在俄国难以得到。因而奥斯特罗格拉茨基和布尼亚科夫斯基几乎成了拉普拉斯主义者，大力宣传拉普拉斯的概率观点，并通过贝叶斯定理计算逆概率。奥斯特罗格拉茨基共发表了 6 篇有关概率论的论文，主要是关于概率论在保险业、养老基金等方面的应用，这些也许得益于他的恩师。总的看来，概率论不是奥斯特罗格拉茨基最感兴趣的课题。然而，布尼亚科夫斯基对概率论很感兴趣，他花费了大量时间来研究概率论。他还特别爱好法语，其多数论文是用法语发表的，因而有人称其为法国数学学派的俄国代表。布尼亚科夫斯基和奥斯特罗格拉茨基都致力于解释各种随机现象，如抽签问题，体育比赛中的机会问题以及保险业和养老基金问题等，这不仅满足了公众的需要，而且也为实际计算提供了正确的基础。我们从中可以窥测到法国数学文化对俄国数学在理论和实际应用中的深刻影响。

由于当时没有俄语的概率论专题研究，更没有概率论的书籍。故布尼亚科夫斯基的专著《数学概率论基础》满足了时代的需求，以至多年为俄国的标准教科书。书中布尼亚科夫斯基继承和发展了拉普拉斯的概率思想，力求阐明和简化拉普拉斯和泊松的有关证明，描述了泊松-孔多塞-拉普拉斯概率专题，并将概率论应用于选举理论，同时注意发展和提炼俄语的专业术语。出于对拉普拉斯的尊敬和爱戴，布尼亚科夫斯基称其专著为俄语版的《分析概率论》。正是这本书使布尼亚科夫斯基重新获得圣彼得堡大学的教授职位，并从 1858 年起成为政府在统计、保险业和养老基金等方面的专家。布尼亚科夫斯基的概率专著在法国也产生了一定的影响，法国数学家比埃奈梅为了读懂其作品而开始学习俄语。曾被选为圣彼得堡科学院通讯成员的法国数学家拉梅，在 1820~1830 年俄国工作期间与布尼亚

科夫斯基和奥斯特罗格拉茨基就概率问题进行了广泛的探讨，对他们在概率方面所取得成果表示敬佩。

2.莫斯科大学的切比雪夫

切比雪夫于 1821 年 5 月 16 日生于俄国奥卡多沃一个贵族家庭。儿时母亲教他读书和写作。有位表姐教他唱歌、学法语和做算术，切比雪夫一直珍藏着其照片直至逝世。1832 年，举家迁往莫斯科。父母聘请了当时莫斯科最有名的家庭教师波戈列日斯基。切比雪夫从波戈列日斯基身上学到了不少知识，并由此对数学产生了强烈的兴趣。

1837 年，切比雪夫考入莫斯科大学，成为哲学系数学物理专业的学生。热努夫(N.E.Zernov，1804~1862)和布拉斯曼(N.D.Brashman，1796~1866)两位老师对切比雪夫的影响较大。热努夫是俄国第一位数学博士，他教切比雪夫“纯数学”理论，使其奠定了坚实的数学基础。在 1841 年 6 月 17 日莫斯科大学的纪念仪式上，布拉斯曼教授作了题为“数学科学对人类智力发展的影响”的演讲，其中他悲观地认为在俄国无论是大学还是研究机构都对概率论这门学科重视不够，并强调指出保险业需要概率论这样严密的数学理论。当时切比雪夫已经完成了大学课程，正在准备硕士考试，听了布拉斯曼的演讲，对他产生了很大的刺激，并暗下决心来研究概率论这门学问。热努夫曾著有一本 85 页的《概率论》小册子，其内容与布尼亚科夫斯基的著作相近，由于当时莫斯科不是俄国的文化中心，因而在俄国产生的影响不大，但对切比雪夫却产生了很大的影响。正是上述种种原因，切比雪夫对概率论开始了研究，并选题《试论概率论中的基础分析》做硕士论文。1844 年 10 月 17 日，切比雪夫的硕士论文定稿。

切比雪夫给出其研究动机：“以初等数学知识证明概率论中的基本定理和主要应用原理；在观测数据的基础上，为各个科学分支提供支持和服务。尊敬的斯特格努夫(C.S.G.Stroganov，1794~1882)阁下使我意识到选择这个题目的重要性……”这篇硕士论文承认没有指导教师。在当时的情况下，切比雪夫可能与热努夫讨论有关概率论问题，但有些概率史家认为是布拉什曼启发了切比雪夫。另外，从通常的教学意义看，斯特格努夫不应是切比雪夫的指导教师。他是一个贵族和能干的官员，在 1835~1847 年负责管理莫斯科地区的教育工作。有人曾这样描写斯特格努夫的功绩“因为他注重提高教师的地位，支持和保护人才，当代人把这些年看作莫斯科大学的黄金时代。他最关心的是学生的利益，对学校设施也做了很大的改进。”因此，切比雪夫来到莫斯科大学是很幸运的，他深受斯特格努夫的赏识，这其中也许是由于布拉什曼的极力推荐。

切比雪夫的硕士论文几乎全是理论性的。他借助 $\ln(1+x)$ 的麦克劳林展开式，对伯努利大数定律作了精细的研究和严格的证明，并指导读者应用标准正态分布表来计算有关概率。这篇文章在当时没有产生很大的影响力，然而在此基础上切比雪夫撰写出一篇优秀的论文《概率论中基本定理的初等证明》，发表在克雷尔(A.L.Crelle，1780~1855)《纯粹与应用数学杂志》上。文中给出了泊松大数定律的证明。由于法国数学家认为概率论还是一个有争议的课题，故在当时没有得到应有的重视。从现在来看，切比雪夫的这两篇论文都是在概率论的新方向——极限理论上迈出了重要的一步。

3.布尼亚科夫斯基的影响

1846 年 7 月，切比雪夫为了寻找一份合适的教学工作来到圣彼得堡大学。1847 年 9 月切比雪夫开始讲授高等代数和数论课程。他的数学才干很快得到布尼亚科夫斯基和奥斯

特罗格拉茨基的青睐。特别是布尼亚科夫斯基对切比雪夫倍加呵护，推荐切比雪夫与他共同整理欧拉的《数论》。切比雪夫从欧拉的著作中体会到了深邃的思想和灵活的技巧结合的魅力，吸引他从事数论的研究。切比雪夫的博士论文《论同余式》就是数论方向的，于1849年5月27日在圣彼得堡大学通过了答辩，并获得圣彼得堡科学院的最高数学荣誉奖。布尼亚科夫斯基从1850年到1859年退休一直讲授概率论这门课程。已晋升为教授的切比雪夫接替布尼亚科夫斯基讲授概率论。这样切比雪夫的研究兴趣再次聚焦在概率论上。退休后的布尼亚科夫斯基仍在大力支持切比雪夫，推荐他到圣彼得堡科学院任职，承担起他与西方数学家的联系工作。布尼亚科夫斯基的法语非常好，这也深深影响了切比雪夫。切比雪夫从表姐那里所学的法语在这里得到进一步加强。不久切比雪夫就完全掌握了法语，他承认是以法语思维来进行科学思考的，而后再以俄文写出来。理论联系实际是布尼亚科夫斯基和切比雪夫的又一个共同爱好所在。他们对机械制作都有着浓厚的兴趣，由于切比雪夫左脚生来有残疾，儿时当其他孩子玩耍时他就静静的做小玩具。布尼亚科夫斯基和切比雪夫共同设计了40余种机器及80余种这些机器的变种，其中有可以模仿动物行走的步行机；有可以自行变换船桨入水和出水角度的划船机；有绘出大圆弧的曲线规等。

4.法国数学文化的熏陶

1836 年，法国数学家刘维尔(J.Liouville，1809~1882)创办了《纯粹数学与应用数学杂志》。他自创办之日起连任主编40载，出版了39期杂志。不少数学家受益于该刊，并从这里踏上数学创作之路，切比雪夫就是其中之一。据刘维尔记载，切比雪夫于 1852 年 6 月~11 月第一次到法国考察，其住处与刘维尔的住所很近。据说切比雪夫自己在房间煮茶，并抱怨周围噪音太大。后切比雪夫又分别与 1856 年、1864 年、1873 年、1875 年和 1878 年出国考察或进行学术交流。1852 年 10 月 12 日，切比雪夫第一次见到比埃奈梅。比埃奈梅于 1796 年 8 月生于巴黎，其父是一个高级官员。1820 年，比埃奈梅进入法国政界，1836 年晋升为一个法国高级检察官。1848 年始比埃奈梅就从事科学研究，曾一度担任统计协会大奖的评委达 23 年之久。他的好朋友有沙勒(Chasles，1793~1880)、拉梅和几何学家凯特尔(L.A.J.Queteler，1796~1874)等。1852 年 4 月 11 日辞职并于 4 月 30 日离开政坛。拉普拉斯的《分析概率论》可谓是比埃奈梅的指路明灯，他的大量工作是解释和推广拉普拉斯的概率理论。出于对拉普拉斯的忠诚和热爱，他曾多次卷入与他人的争论中，其对手主要是泊松、柯西和贝特朗(J.Bertrand)等。可惜的是，由于比埃奈梅不停颤抖的手及难以入睡的毛病，阻碍了他进行深入思考和研究，以致没有任何弟子，也没有留下任何专著。

与切比雪夫见面时，比埃奈梅已完全退出政坛。尽管他们相差 25 岁之多且比埃奈梅身体状况一直较差，但他们成为忘年交，共同探讨数学问题尤其是有关概率论的研究课题。切比雪夫的法国好朋友还有埃尔米特(Charles Hermite， 1822~1901)和刘维尔，他们年龄相仿，志趣相投。正是在埃尔米特和刘维尔的大力宣传和帮助下，切比雪夫很快在国际数学界有了一席之地。1855 年，切比雪夫以俄语发表了一篇论文，他在正交化多项式的基础上，将设置的矩阵正交化，达到了柯西最小二乘法的最初目的。比埃奈梅于 1858 年将这篇论文翻译成法语发表，在所给的导言中提及与柯西 1853 年关于插值法的一场激烈争论。

由于比埃奈梅的干涉，切比雪夫表明其简洁数学计算方法与通常非正交化的最小二乘法是相当的，进而与柯西的插值法是可以比拟的。众所周知的切比雪夫不等式，确切来说应为比埃奈梅-切比雪夫不等式。1867 年，切比雪夫将这篇论文同时以俄语刊登在圣彼得

堡和以法语发表在《刘维尔杂志》上。直到发表后，切比雪夫才知道比埃奈梅早在 1853 年就给出了相关证明。刘维尔将比埃奈梅的论文刊登在切比雪夫的论文前面，并给出一个编者按，暗示这两篇文章的相互联系。也许比埃奈梅认为这个不等式与他的线性最小二乘法主题及维护拉普拉斯的观点相比是微不足道的，因而没有和切比雪夫沟通。而切比雪夫充分认识到这个不等式的意义。他们当时所给的不等式如下

$$P(|s_N - Es_N| \geq \varepsilon) \leq \frac{\sum_{i=1}^{N} VarX_i}{\varepsilon^2},$$

这里 $X_1, X_2, \cdots, X_N$ 是相互独立的，$s_N = \sum_{i=1}^{N} X_i$。

显然其实质上为弱大数定律的一般表达式。

切比雪夫所证明的是不同分布的离散型随机变量序列，由于这些限制其证明是比较复杂的，比埃奈梅所证明的是独立同分布的随机变量序列。但从本质看，证明思路是相同的。该不等式的简洁证明方法已为当今教科书的内容，这里不再赘述。

1874 年，切比雪夫在递交给法国学术会议的一篇论文中，明确承认了比埃奈梅对该不等式的优先发明权。这篇论文也被《刘维尔杂志》刊出。切比雪夫写道："我所给出的简洁而严格的伯努利大数定律的证明，实际上是比埃奈梅先生方法的一个演绎结果。在他所给的结论中，很容易导出伯努利大数定律"。

在比埃奈梅的证明中含有所谓的"矩方法"的思想，利用矩方法可以解决许多困难的极限估计问题，直到今天该法仍被广泛应用。切比雪夫一定花费不少时间来研究其证明。在切比雪夫 1874 年的论文中写道："杰出的专家们给出一个好的方法。这个方法是以

$$\int_0^A f(x)dx, \int_0^A xf(x)dx, \int_0^A x^2 f(x)dx, \cdots,$$

来确定积分值

$$\int_0^a f(x)dx,$$

这里 $A > a$，且 $f(x)$ 是未知函数并假定在积分区间内恒为正值"。

1887 年，切比雪夫以"矩方法"给出了独立但不同分布随机变量序列之和的中心极限定理。这种方法被其学生马尔可夫(Markov，1856~1922)继承并推广。

5.切比雪夫大数定律的创立

第一个大数定律是雅各布·伯努利提出的。因其遗著《猜度术》于 1713 年出版，故概率史家称 1713 年为伯努利大数定律创立年。伯努利大数定律给出了频率估计概率的理论依据，同时开创了概率论中极限理论的先河。1837 年，泊松对大数定律提出一个较宽松的条件，进而得到泊松大数定律。之后，由于有些数学家过分强调概率论在伦理学中的应用甚至以此来阐明隐蔽着的神的秩序，又加上概率论自身基础不牢固，大多数数学家往往把概率论看作有争议的课题，排除在精密科学之外。切比雪夫正是在概率论门庭冷落的年代从事其研究的。由上所述，切比雪夫的数学研究深受布尼亚科夫斯基、奥斯特罗格拉茨基老一辈俄国数学家和比埃奈梅、埃尔米特等法国数学家的影响。

19 世纪后期，极限理论的发展成为概率论研究的中心课题。1866 年俄国数学家切比

雪夫发表的论文《论均值》在这方面迈出了决定性的一步。

切比雪夫在该文中从后来以他的名字命名的不等式出发，建立了关于独立随机变量序列的大数定律。切比雪夫大数定律使伯努利定理和泊松大数律都成为其特例。翌年切比雪夫又推广棣莫弗——拉普拉斯极限定理，建立了有关各阶绝对矩一致有界的独立随机变量序列的中心极限定理。切比雪夫的成果后被他的学生马尔可夫和李亚普诺夫等发扬光大，深刻影响了20世纪概率论的发展进程。

在1866年发表的论文《论均值》中，切比雪夫提出了著名的切比雪夫大数定律，该论文给出如下三个定理：

定理1：若以$a,b,c,\cdots$表示量$x,y,z,\cdots$的数学期望，用$a_1,b_1,c_1,\cdots$表示相应的平方$x^2,y^2,z^2,\cdots$的数学期望，则对任何α和$x+y+z+\cdots$，落在

$$a+b+c+\cdots+\alpha\sqrt{a_1+b_1+c_1+\cdots-a^2-b^2-c^2-\cdots}$$

和

$$a+b+c+\cdots-\alpha\sqrt{a_1+b_1+c_1+\cdots-a^2-b^2-c^2-\cdots}$$

之间的概率总大于$1-\dfrac{1}{\alpha^2}$。

如果N是$x,y,z,\cdots$的个数，在我们上面的定理中令

$$\alpha=\frac{\sqrt{N}}{t},$$

并用N除以和$x+y+z+\cdots$及它的上下限

$$a+b+c+\cdots+\alpha\sqrt{a_1+b_1+c_1+\cdots-a^2-b^2-c^2-\cdots}$$

和

$$a+b+c+\cdots-\alpha\sqrt{a_1+b_1+c_1+\cdots-a^2-b^2-c^2-\cdots},$$

我们将得到如下的关于均值的定理。

定理2：若以$a,b,c,\cdots$表示$x,y,z,\cdots$的数学期望，用$a_1,b_1,c_1,\cdots$表示相应的平方$x^2,y^2,z^2,\cdots$的数学期望，则不论t取何值，N个量$x,y,z,\cdots$的算术平均值和它们相应的数学期望的算术平均值的差不超过

$$\frac{1}{t}\sqrt{\frac{a_1+b_1+c_1+\cdots}{N}-\frac{a^2+b^2+c^2+\cdots}{N}}$$

的概率对任何t都将大于$1-\frac{t^2}{N}$。

这实际上就是切比雪夫不等式，它是证明大数定律的工具，在此基础上发展起来的一系列不等式是研究中心极限定理的有力工具。

由于分数$\dfrac{a_1+b_1+c_1+\cdots}{N}$和$\dfrac{a^2+b^2+c^2+\cdots}{N}$表示量$a_1,b_1,c_1,\cdots$和$a^2,b^2,c^2,\cdots$的均值，当数学期望$a,b,c,\cdots$和$a_1,b_1,c_1,\cdots$不超过给定的有限值时，不论$N$多大，表达式

$$\sqrt{\frac{a_1+b_1+c_1+\cdots}{N}-\frac{a^2+b^2+c^2+\cdots}{N}}$$

将是一有限值。因此，通过取t充分大，可使

$$\frac{1}{t}\sqrt{\frac{a_1+b_1+c_1+\cdots}{N}-\frac{a^2+b^2+c^2+\cdots}{N}}$$

充分地小。由于不论t取何值，如果N趋于无穷，则分数$\frac{t^2}{N}$将趋于0。利用第二个定理，我们得出如下结论：

定理3：如果量$U_1,U_2,U_3,\cdots$和它们的平方$U_1^2,U_2^2,U_3^2,\cdots$的数学期望不超过一给定的有限值，则N个这些量的算术平均值和它们数学期望的算术平均值之差不小于某一给定的概率，且当N趋于无穷时，其值趋于1。

这就是切比雪夫大数定律，用今天的符号可表示为：

定理3′：设$X_1,X_2,\cdots,X_n,\cdots$是两两不相关的随机变量序列，且其方差一致有界，则对任意的$\varepsilon>0$，皆有

$$\lim_{n\to\infty}P(|s_N-Es_N|<\varepsilon)=1 ,\tag{2-10}$$

其中$s_N=\sum_{i=1}^{N}X_i$。

其证明由切比雪夫不等式容易得出。若随机试验中的每次试验随机事件发生的概率相等，则为伯努利大数定律，又因相互独立的随机变量列必定两两无关，故泊松大数定律也是切比雪夫大数定律的特例。翌年切比雪夫又推广了棣莫弗——拉普拉斯极限定理，建立了有关各阶绝对矩一致有界的独立随机变量序列的中心极限定理。他还研究了相互独立正态变量和的分布函数的收敛条件，提出其按$n^{-1/2}$方幂渐进展开的猜想(n为独立随机变量的数目)。这一猜想完全被后来的研究所证实。

著名数学家柯尔莫哥洛夫(A.N.Kolmogorov，1903~1987)将切比雪夫的概率思想概括为

“从方法论的观点来看，切比雪夫所带来的根本变革的主要意义不在于他是第一个在极限理论中坚持绝对精确的数学家(棣莫弗、拉普拉斯和泊松的证明与形式逻辑的背景是不协调的，他们不同于雅格布·伯努利，后者用详尽的算术精确性证明了他的极限定理)，其工作的主要意义在于他总是渴望从极限规律中精确地估计任何次试验中的可能偏差并以有效的不等式表达出来。 此外，切比雪夫清楚地预见到诸如随机变量及其期望(平均)值等概念的价值，并将它们加以应用的第一个人。这些概念在他之前就有了，它们可以从事件和概率。这样的基本概念导出，但是随机变量及其期望值是能够带来更合适与更灵活的算法的课题。

“马尔可夫（Markov，1856~1922）是俄国数学家。1856 年 6 月 14 日生于梁赞，1922 年 7 月 20 日卒于圣彼得堡。

马尔可夫上中学时，大部分课程学得不好，唯独数学成绩常常都得满分，并开始自学微积分，有一次他独立地发现了一种常系数线性常微分方程的解法，就写信给著名数学家布尼亚科夫斯基，信被转到圣彼得堡数学系科尔金和佐洛塔廖夫手里，从此马尔可夫与圣彼得堡大学的数学家建立了联系。1874 年考入圣彼得堡大学数学系学习，在学习期间他深受切比雪夫、科尔金、佐洛塔廖夫等数学家的启发和影响，1878 年大学毕业，并以《用连分数求微分方程的积分》一文获金质奖章。1880 年以题目为《论行列式为正的二元二次齐次》的论文取得硕士学位并在圣彼得堡大学任教。1884 年获物理数学博士学位，1886 年成为教授，1890 年当选为圣彼得堡科学院候补院士，1896 年当选为院士。1905 年退休

时圣彼得堡大学授予他功勋教授称号。

马尔可夫研究的范围很广，对概率论、数理统计、数论、函数逼近论、微分方程、数的几何等都有建树。

在概率论方面，他深入研究并发展了其老师切比雪夫的矩方法，使中心极限定理的证明成为可能。他推广了大数定律和中心极限定理的应用范围。他提出并研究了一种能够用数学分析方法研究自然过程的一般图式，这种图式后人即以他的姓氏命名为马尔可夫链。他还开创了一种无后效性随机过程的研究，即在已知当前状态的情况下，过程的未来状态与其过去状态无关，这就是现在大家耳熟能详的马尔可夫过程。马尔可夫的工作极大地丰富了概率论的内容，促使它成为自然科学和技术直接有关的最重要的数学领域之一。在数理统计方面，他还引入了等价互不相容概念和有效性统计原理。

李雅普诺夫（Lyapunov，1857~1918)是俄国数学家、力学家。1857 年 6 月 6 日生于雅罗斯拉夫尔，1918 年 11 月 3 日卒于敖德萨。

李雅普诺夫 1876 年中学毕业时，因成绩优秀获金质奖章，同年考入圣彼得堡大学物理数学系学习，当他听了著名数学家切比雪夫的讲座之后即被其渊博的学识深深吸引，从而转到切比雪夫所在的数学系学习，在切比雪夫、佐洛塔廖夫的影响下，他在大学四年级时就写出具有创见的论文，而获得金质奖章。1880 年大学毕业后留校工作，1892 年获博士学位并成为教授。1893 年起任哈尔科夫大学教授，1900 年初当选为圣彼得堡科学院通讯院士，1901 年又当选为院士，并兼任应用数学学部主席。1909 年当选为意大利国立科学院外籍院士，1916 年当选为巴黎科学院外籍院士。

李雅普诺夫是切比雪夫创立的圣彼得堡学派的杰出代表，他的建树涉及多个领域，尤以概率论、微分方程和数学物理最有名。

在数学中以他的姓氏命名的有：李雅普诺夫第一方法，李雅普诺夫第二方法，李雅普诺夫定理，李雅普诺夫函数，李雅普诺夫变换，李雅普诺夫曲线，李雅普诺夫曲面，李雅普诺夫球面，李雅普诺夫数，李雅普诺夫随机函数，李雅普诺夫随机算子，李雅普诺夫特征指数，李雅普诺夫维数，李雅普诺夫系统，李雅普诺夫分式，李雅普诺夫稳定性等，而其中以他的姓氏命名的定理、条件有多种。

在概率论中，他创立了特征函数法，实现了概率论极限定理在研究方法上的突破，这个方法的特点在于能保留随机变量分布规律的全部信息，提供了特征函数的收敛性质与分布函数的收敛性质之间的一一对应关系，给出了比切比雪夫、马尔可夫关于中心极限定理更简单而严密的证明，他还利用这一定理第一次科学地解释了为什么实际中遇到的许多随机变量近似服从正态分布。他对概率论的建树主要发表在其 1900 年的《概率论的一个定理》和 1901 年的《概率论极限定理的新形式》论文中。他的方法已在现代概率论中得到广泛的应用。

第 3 章　现代概率论——测度概率论

概率论在物理、生物等领域的应用，也需要对概率论的概念、原理做出解释。于是，1900 年著名数学家希尔伯特在巴黎召开的第二次国际数学家大会上作的给出了对 20 世纪数学的进程有重大影响的二十三个“数学问题”著名的报告中的第 6 个问题：“对几何基础的探讨提示我们，根据这个模式，我们应当公理化地对待这些物理的法则，其中数学已经起了显著的作用：其中第一个也是最重要的一个是概率演算和关于数学概率论的公理化，我们渴望的似乎是任何逻辑的研究与数学物理，特别是气体力学理论中的均值方法的严格与相容的发展同步进行”，就呼吁把概率论公理化。因此，这一问题成为当时数学及整个自然科学的最迫切的问题之一。1933 年以后，概率论的研究主要采用测度论方法，这一时期称为测度概率论时期，这时概率论公理化，开创了现代概率论的新时期。

3.1 拉普拉斯概率理论的衰落

在概率论历史上，拉普拉斯被认为是古典概率论的集大成者，其皇皇巨著《分析概率论》在 19 世纪概率论发展中的作用可以与欧几里得的《几何原本》在几何学中的作用相媲美，它左右了 19 世纪概率论的发展。然而，伴随着拉普拉斯概率论的普及和发展，19 世纪 30 年代以来对拉普拉斯概率理论的争论和批评的声音也一直没有间断过。在 20 世纪初期，作为概率论历史上一个重要发展阶段的拉普拉斯的概率论终于被公理化的概率论所代替。拉普拉斯概率理论的发展史是许多数学理论发展的一个缩影，它可以为我们从哲学的角度研究数学理论的发展模式提供一个典型案例。

1.关于概率论在法律判决中应用的争论

拉普拉斯的概率理论是启蒙运动时期思想潮流的一种具体尝试，这种尝试试图把人类的所有行为都聚集在理性的法则之下，它秉承了从 17 世纪雅各布•伯努利开始的对于概率的传统理解，即概率不仅可以描述人类理性而且也可以将理性推理系统化。一旦这个理想得以实现，那么，这些结果能够超越受过良好教育的人群而推广至更广泛的人群，从而人们获得在道德、经济、法律等所有领域中有教育价值的社会共识。正是凭借这种以概率规则引导人类理性的信念，拉普拉斯把概率论应用于当时的各个领域中去，例如，天文地理、人口统计、赌博输赢、人寿保险、法庭判决等。但是，18 和 19 世纪发生的一系列社会活动和事件为数学家们的努力打上了问号。例如，在 18 世纪，许多欧洲国家各种彩票发行盛行，很少有人理会数学家或者哲学家们对于“赌博是非理性的”呼吁和研究，更不关心这些社会精英们对于理性价值的数学计算。所以，对于那些希望鼓励人们由数学推理引导行为的人来说，赌博活动就成为某种类似悖论的东西，当数学如此清楚地显示了赌博的非理性，为什么那么多人义无反顾地参与到这些活动中去?至于概率在法庭判决中的应用，直到 19 世纪 30 年代，法官们对于用概率方法研究的法律结果几乎没有一点反响。所以，尽管拉普拉斯和他的后继者们关于统计的稳定性的富有煽动性的语言激发了人们的想象

力和激情，然而，由于拉普拉斯、泊松、凯特勒等人对于概率论的作用不加限制的夸大，并常常在证据不充足的情况下做出结论性的宣称，这些行为常常使人怀疑概率统计是否是一门自我吹嘘、玩弄笔墨、轻率肤浅的知识。这种状况为拉普拉斯的概率论投下了一丝阴影。1835 年，拉普拉斯概率论的直接继承者、法国数学家泊松(S.D.Poisson)在他的论文《关于犯罪和民事判决的概率的研究》中仍然宣称“数学的严格性可以证明判决结论是否正确”。这种不顾现实的傲慢在他的科学院同事中激起了强烈的反映，由此引起一场激烈的争论。这场争论的参与者主要有法国科学院的著名科学家纳瓦尔(Claude Navier)、普安索(Louis Poinsot)和迪潘(Charles Dupin)等人。

最先对泊松提出批评的是普安索，他说：

“概率演算在法庭中的应用触犯了智力这个神圣的字眼，例如，用一个数表示证言的正确性，这样就把人比作骰子，其中每一个有 n 面，有的面对应着错误，有的面对应着正确，并以同样的方式处理其他道德问题，把它们转换成如此众多数字的分数。还竟敢在这种计算的最后得出结论以引导理性的人们对一个犯罪案件作出判断，或者作出一个决定，或者给他人提出重要问题的建议——这些对我来说是一种智力的堕落，是科学的虚假的应用，是绝对不可信的”。

泊松在 1836 年 4 月 18 日的一次科学院会议上回应了这些人的批评。他指责普安索等人对于拉普拉斯的判决概率的曲解。泊松特别提醒人们注意由司法统计局报告的犯罪率惊人的规则性，他再次明确了对大数定律的普适性的信心——任何现象的足够大的集合，不管是道德的还是物理的现象——最终都将展示出规则性。

关于这场争论，一位名叫古罗(Charles Gouraud)的法国学者在 1848 年提交给道德和政治科学学院的一篇名为《确定性的理论》的文章中作了生动地描述。他说，普安索和迪潘之所以站出来与泊松论战，除了学术上的看法的不一致外，还有其他的原因。迪潘和普安索是几何学家蒙日的狂热追随者，他们热衷于几何方法，他们经常介于法国分析学派(泊松是最有名的代表人物之一)和蒙日的几何学派之间的争论之中。所以，在泊松的著作出版以后，乘借着法国几何学派对分析学派展开论战的氛围，以泊松的工作在科学院引起的争论为靶子，几何学派的信徒们的怒气全部朝向泊松并不令人惊奇。

正如古罗所说，从表面上来看，这是一场把概率演算应用到“道德世界”中是否合适的争论，而实质上则是对以拉普拉斯为代表的整个概率论这一学科的性质和基础的反思。不过这些哲学家们鉴于拉普拉斯的威望和地位，此时还没有直接将矛头对准拉普拉斯，然而这场争论却成为进一步抨击拉普拉斯概率理论的一个导火索。

2.柯西的“新数学”的兴起及其对于最小二乘法的责难

哲学家们对于概率论的反击是出于担心数学侵入道德领域而使之失去了自己的本色和传统而变得面目全非，他们反感的是数学的傲慢。这些批评是外在的，给人以隔靴搔痒之感，所以对于以拉普拉斯权威为核心的概率论的地位并未有明显的动摇。在这场争论的开始数学家们并未参与其中，他们只是骑墙观望，但是，这场争论后来的发展也将数学家卷入了其中。来自数学家内部的批评显然比哲学家的指责更能够击中拉普拉斯概率论的缺陷和要害。在某种程度上，数学界参与这场讨论反映了在数学发展中，在惯例或制度水平上以拉普拉斯为代表的 18 世纪的科学院所实践的数学与以蒙日、柯西为代表的类似于多科工艺学院的高等教育机构中所教授的数学的分歧，并进而反映了在 18 世纪肩负重任的

数学(当时的唯一标准是数学应用的有效性和潜在的可实用性)与在高等教育机构中所从事的数学(以科学的教学法的要求作指导，转向于以严格性和简单性为标准)之间的决裂。从 19 世纪初期开始，数学界正在经历着一场深刻的变革，对以往的各个数学理论分支、研究方法等展开了全面的检讨，以重建一种“新的数学”。在这场运动中，柯西是最主要的代表人物之一。拉普拉斯的概率论深深地扎根在 18 世纪的数学传统中，他对概率论这门学科的理解，以及他对概率论研究的实践和方法等都带有浓重的 18 世纪数学的风格特征，他所理解的概率论就是一门自然科学，是一门应用学科，检验它的价值的重要标准是它在实践中的有效应用， 而不是其自身的严格和逻辑上的相容。所以，在这场运动中，对他的概率论的检查和批判也就日益深入地纳入了数学家们的日程。

在巴黎多科工艺学院，柯西的讲义《分析教程》(1821)在当时是作为数学的中心内容——分析学的一个导引，这个教程成为 19 世纪所出现的新数学的一个标志。在这部著作的序言部分，柯西特别强调：尽管他的工作是建立在 18 世纪的数学结果的基础上，但他的工作在所应用的严格程度上，以及数学应用的范围被严格限制的方面不同于他的前辈们。柯西明确宣称，这部教程将主要侧重于两方面的讨论：一方面，将尽力使数学分析完善，另一方面，尽力避免“数学万能” 的宣称。

柯西倡议对数学的应用必须加以限制：

“让我们热烈地在数学领域的内部探寻数学科学的研究，而不必急于开拓它们的领地，并且不要想通过数学公式研究历史，或者用代数定理或微积分批准和认可道德条例。”

在这里柯西明显地把自己对于数学应用的看法区别于拉普拉斯的观点：

“让我们将应用于政治和道德科学中的方法建立在观察和演算之上，这些方法曾经如此成功地应用于自然科学，让我们不要提供哪怕最少无效的以及常常有危险的阻力于知识的不可避免的进步。”

柯西还强调，这些讨论不仅仅是数学应用方面的限制问题，而是涉及对所有数学进行改革的方向问题。他说，限制的方法，如指明函数的定义域和值域、级数收敛的区域、更一般的数学命题的有效性区域等将把数学从 18 世纪所宣称的普遍有效性的困境中解救出来。柯西后来把这种限制应用到了拉普拉斯的概率理论的中心之一：最小二乘法。柯西首先改革的是拉普拉斯关于定律的含混表达方式：拉普拉斯在他的《分析概率论》中指出：“不管误差的频率定律，一种依赖于相对频率 k 是什么，必须说明这种方法(最小二乘法)是应当优先选择的。”柯西认为，一个数学命题必须以两种方式给出：命题要么能够被证明，要么作为不加证明的原始命题。这里 k 主要等同于误差的变量，而拉普拉斯的公式中暗示了所有变量的存在性，但是这个断语不能够被证明，所以，为了符合数学命题的条件，必须把拉普拉斯的默认的一些结论转化成一个正确的断言。其次，柯西也否定了这个结果的普遍有效性，柯西于 1853 年成功地找到一个最小二乘法无效的例子，这就是后来以他的名字命名的分布——柯西分布。

由此，柯西认为他已成功地说明了“由最小二乘法得到的未知数的值并非是感觉上最可能的值。”柯西对最小二乘法的质疑对于在数学界掀起对概率论的反思运动具有极大的号召力，由于柯西当时已是法国科学界资格最老的人，尽管在他的《分析教程》中没有提到拉普拉斯的名字，但是也为以后数学界直接挑战拉普拉斯的概率理论提供了依据。

3.对于拉普拉斯概率理论的基础的批评

19 世纪末期，对于概率论的讨论开始集中在拉普拉斯关于“概率”的定义上。早在1774 年的“论概率”一文中，拉普拉斯就给出了现在所称的“概率的古典定义”。在《分析概率论》中拉普拉斯是把这个概念作为他的概率理论的出发点的。拉普拉斯的概率定义主要是建立在他的两个著名的原则之上，其中一个原则是“贝叶斯法则”，另一个是“不充足理由律”。在两种情况下都涉及先验概率和后验概率的问题，这两个原则相互配合，使得拉普拉斯得以把逆概率广泛应用于机会的游戏、观察的误差，男女出生的比率，以及法庭的判决等方面。拉普拉斯对于这两个原则的应用带有明显的主观倾向，他自己也非常清楚这一点，例如他在“由无知所导致的等可能性”中，区分了假设的“相对可能性”和事件自身的本性所导致的“绝对可能性”：绝对可能性对应着事件的恒定原因，相对可能性只是与我们关于这些恒定原因的知识状态有关。拉普拉斯一再警告混淆绝对和相对可能性可能会导致数学上的不相容，所以根据绝对可能性的新知识，概率必须不断地改正对相对可能性的估计：既利用逆概率公式，根据已观察到的结果计算原因的概率，然后再根据给定的已发生的事件的记录等信息计算将来事件的概率。

但是，在实践中，拉普拉斯几乎总是假设所有的情形是先验地等可能的。这是引起后人把矛头直接指向拉普拉斯的概率理论的切入口，也被认为是拉普拉斯的概率理论的阿基里斯的脚后跟，对于它的攻击是致命的，它暴露了拉普拉斯的整个概率论体系的基础的薄弱。19 世纪后期，有许多哲学家和数学家参与了这个问题的讨论。总的来说，对于拉普拉斯的概率的批评主要有以下几个方面：

(1)拉普拉斯的概率定义是循环的。米泽斯首先对此提出批评，他说：“定义中的等可能性”完全等同于“等概率性”，所以，这个定义是“恶性循环的”， 在逻辑上是不相容的。

(2)对于“等可能性”的定义依赖于拉普拉斯再三强调的“没有理由使人们相信一些事件的一个应当比其他的更倾向于发生”这种所谓的“不充足理由律”，而这个原则本身是不自恰的。因为它首先是只适用于对目前情形没有经验的人； 其次，没有一个是否可以决定它被正确应用的标准，在实践中很难应用它，即使一些著名的数学家也难以把握，如怎样判断等可能的情况?

(3)拉普拉斯的概率定义只适用于非常有限的一类问题。这样得出的概率的值只是有理数，对于连续性的问题，拉普拉斯的概率定义是无能为力的。所以为了确定古典概率中的各种事件的概率，必须引入某些无限的测度，如长度或面积等，否则就容易产生混乱。

拉普拉斯概率与生俱来的缺陷在德国哲学家克莱斯(Von Kries)和法国数学家贝特朗(JBertrand)提出的悖论中得到了淋漓尽致的反映。这个悖论的数学语言可以表述如下：在半径为r的圆C内“随机”地作一个弦AB，问弦长l大于此圆的内接正三角形的边长$\sqrt{3}r$的概率等于什么?贝特朗给出三种解法，分别得出的结果是

$$p=\frac{1}{4},p=\frac{1}{3},p=\frac{1}{2}$$

这样以三种看起来推理都正确的方式得到三个不同的值，不难想象发现者当时的困惑，他们称这个问题是“不详的征兆”。由于“概率”这个概念是概率理论展开的最基本的出发点，就像集合论中的“集合”概念，所以这个悖论的提出所引起的恐慌并不亚于在

20 世纪初集合论中出现罗素悖论时所引起的恐慌。

3.2 概率论的公理化历程

一直被拉普拉斯等人当作应用数学的概率论在他的继承者泊松和凯特勒之后却丧失了早期最吸引人的领域——在社会和道德领域中的应用。然而这种厄运并不是终结，由于柯西的新数学的兴起，以及对拉普拉斯概率论中的不相容性的批评的增多，19 世纪后期，拉普拉斯的概率理论面临着被驱逐出一门学科的境地。的确，在泊松之后，除了教科书的作者们，很难发现一个年轻的数学家认真地从事拉普拉斯的概率理论研究，这种状况引起了人们对于拉普拉斯概率理论的反思性的评论，有人甚至认为拉普拉斯并不是促进了 19 世纪概率论的发展，因为他没有指出概率论进一步发展的新方向，也没有提出概率论应用的新的领域，而是把概率论带入了死胡同。

为了挽救陷于不能自拔的境地的概率论，在 1900 年的第二次国际数学家大会上，著名数学家希尔伯特在著名的演讲“数学问题”中提出：“对几何基础的探讨提示我们：根据这个模式，我们应当公理化地对待这些物理的法则，其中数学已经起了显著的作用：第一个也是最重要的一个是概率演算和关于数学概率论的公理化”。(第 6 个问题)。实际上，希尔伯特的愿望正在实现，在 19 世纪末和 20 世纪初期，概率论中发生了一些重大的变化，一方面，新的应用领域的开拓为概率论在 20 世纪的发展继续提供了强大的动力，例如孟德尔神父(G.Mendel)在 1865 年把概率论与遗传学和杂交配种联系起来；高尔顿(F.Galton) 在他的优生学的研究中于 1877 年发现了回归和相关系数法则；1894 年始，皮尔逊(K.Pearson)把概率应用于生物学并创立了我们今天称为生物统计学的学科领域等。另一方面，在数学的内部，柯西和希尔伯特等人所提倡的新的数学标准——数学的公理化和严格性为概率论注入了新的活力和发展契机。

数学的公理化和严密化趋势起源于 17、18 世纪以来所产生的一些新的数学学科缺乏严密的基础。对于概率论的公理化的思考和尝试开始于 19 世纪中叶。较早思考这方面的人是布尔(G.Boole)，布尔发现他自己发明的符号代数的思想可以为混乱不堪的拉普拉斯概率理论提供一个严密的基础，这种符号代数的主要方法就是公理的方法。

他宣称：

“概率论被列入纯粹数学之中的宣称必须建基于这种程度上，其中它满足以下的条件：

(1)赖于建立其方法的法则应当有公理的本质。

(2)它们应当导致能够精确证明的结果，而证明是可能的。

(3)它们应当能够导致一个与其所有的部分和程序都相容的系统的发展。除了这些存在于事物本质的东西之外，不能承认也不能强加任何限制。”

这些论述已蕴涵了明确的公理化的思想。然而，布尔并未完成他的计划，他没有从混乱中清理出一个清晰的公理系统。

20 世纪前，概率论已经有了很多重要的结果，但当时概率论方面的论文和著作除少数外都明显缺乏数学的严格性。甚至连庞加莱这样的大家也不能把概率论演绎成一门逻辑上完美的数学学科。这是因为 19 世纪的分析本身就没有严格化，以他为研究工具的概率论的严格化就可想而知了。虽然后来分析的基础严密了，但概率论公理化的测度论还未发明，

因此，概率论的不严密是不可避免的。

概率论中著名的“贝特朗悖论”就是在这时候出现的。1889 年贝特朗在他的《概率论》一书中给出了这一问题：在半径为 1 的圆内随机地取一条弦，问其长超过该圆内接等边三角形的边长的概率为多少？贝特朗给出了三种不同的解法。

解法 1：任何弦交圆周两点。不失一般性，先固定其中一点于圆周上，以此点为顶点作一内接等边三角形。显然只有落入此三角形的弦才满足要求，而这种弦的长度为整个圆周的 1/3，故所求概率为 1/3(图 3–1)。

解法 2：弦被其中点唯一确定，当且仅当其中点属于半径为 1/2 的同心圆时，弦长大于内接等边三角形边长，而此小圆面积为大圆面积的 1/4，故所求概率为 1/4(图 3–2)。

解法 3：弦长只跟它与圆心的距离有关，而与方向无关，因此可假定它垂直于某一直径。对于这种弦，当且仅当它与圆心的距离小于 1/2 时，其长才大于内接等边三角形的边长。因此所求概率为 1/2(图 3–3)。

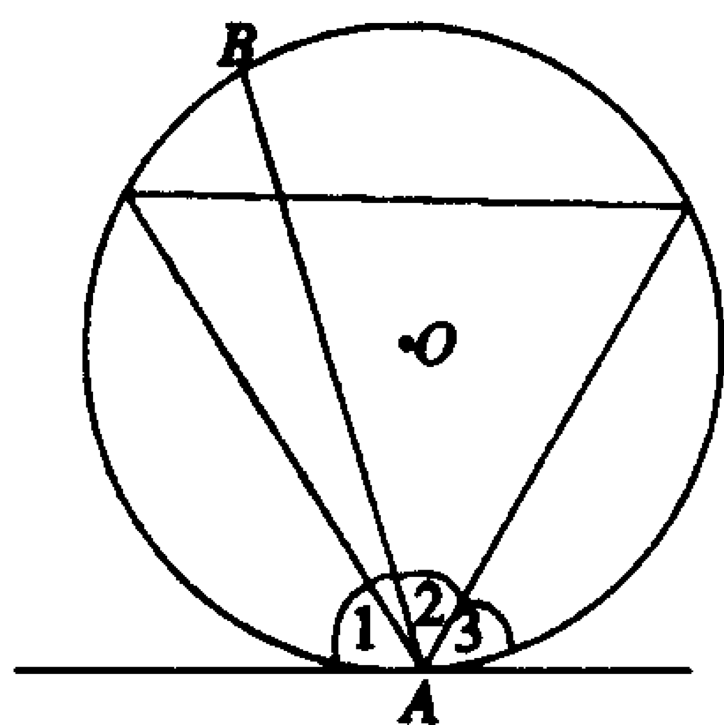


图 3–1

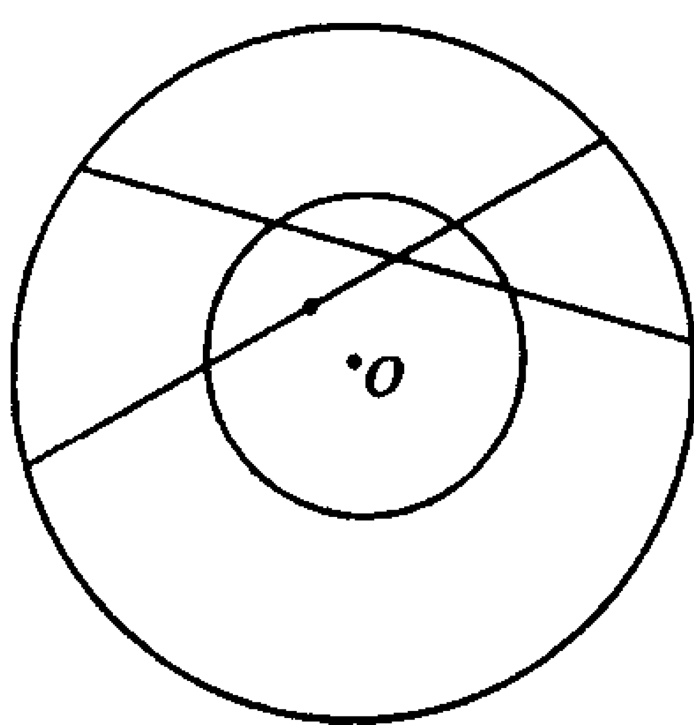


图 3–2

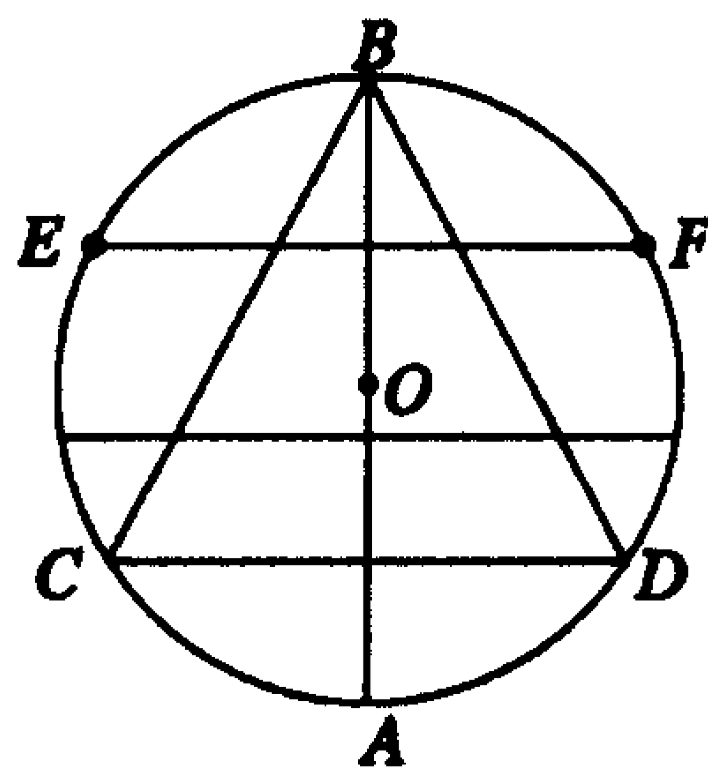


图 3–3

一个问题却得到三个不同的结果，这似乎与数学的一个显著特点——数学论断的确定不变性相矛盾。庞加莱指出悖论的根源在于，无论三种情形下的哪一种，都假定各自的参数均匀地分布在给定的区域里。即，解法 1 中，假定一端固定而另一端点在圆周上均匀分布；解法 2 中，又假定弦的中点在圆内均匀分布；而解法 3 中，假定弦的中点在直径上均匀分布；因此事实上三个问题都被解出。

由于原题中“随机地取一条弦”的提法太不明确了，不同的解释反映的随机事件不同，那得到的结果也是不同的，便很自然了，相对于每种解释的结果都是正确的。

同一时期还出现了许多悖论，正是这些问题促使人们思考概率论的基础问题。其中尤为突出的是人们对概率这一基本概念的认识。历史上曾出现了概率的各种定义，如古典定义、几何定义、频率定义等，每个都有其适用的地方，也自有其局限性。以拉普拉斯为代表提出的古典定义，必须先要确定等概率的初等事件。在类似投掷骰子和暗匣中摸球这类问题中，人们是可以接受等概率的判断的。在另一类问题中，例如明天是否下雨，去判断哪些是等概率事件就不容易了。另外，频率学派(也叫经典学派)对概率的理解就是频率的稳定性，离开了重复试验，就谈不上如何去理解概率。这些概率局限于实在的物质世界，称之为实体概率。

但必须还得承认有另一种概率，它存在于人们的主观世界内，反映了人们对某些事物的一种信任的程度，是对事务的不确定性的一种主观判断，是与个人的、心理上的……各种因素有关的，称之为主观概率。1921 年，凯恩斯首先提出，概率是根据某些命题的知识对一个命题作出的一种有理由的信念。对这种信念无法给出数值，只能进行比较。但是他

没有提供合理的规则来计算概率，甚至如何比较也没有规则。后来 1931 年拉姆齐提倡用打赌的思想来取得每个人的概率。

凯恩斯（Keynes，1883~1946）是英国数学家、经济学家。1883 年 6 月 5 日生于剑桥，1946 年 4 月 21 日卒于苏萨克斯的费尔。

凯恩斯 1905 年毕业于剑桥大学，1906 年以第二名的成绩考入国家行政机关，不久去印度公司服务。两年后回到母校剑桥大学参加经济学系工作，同时还受聘于皇家委员会，协理有关印度方面的财经与流通问题。1913 年出版《印度的通货和财政》一书，而崭露头角。1915 年任职于英国财政部。1919 年作为英方的主要代表出席了在巴黎举行的国际和平会议。凯恩斯曾任英国上议院议员，他是英国皇家学会会员和其他几个科学文化机构的成员。

凯恩斯在数学上的贡献是概率论。凯恩斯是主观概率学派的代表，他主张把任何命题都看做是事件。例如，“明天将下雪”，“学校里有教师”，“张三将死”等。他把一事件的概率看做是人们根据经验对该事件的可信程度，而与随机试验没有直接联系，因此，通常称为主观概率。主观概率的最大影响不在概率论领域自身，而在数理统计中出现的贝叶斯统计学派。和主观概率派相对立的是以米泽斯为代表的频率理论学派。凯恩斯 1911 年著有《概率论》，但 1921 年才正式出版。该书的特点是采用了许多现代的数学符号，并试图为概率论建立一个坚实的数学基础。

凯恩斯在经济学方面深入探讨了特殊商品的供需，各种生产原料的分配、收入的分配等问题。他对 20 世纪 20 年代资本主义世界经济萧条有深入的研究。他 1936 年出版的《就业、利息和货币通论》是经济学领域里的一本名著。

最早对概率论严格化做出尝试的是俄国数学家伯恩斯坦和奥地利数学家米泽斯。他们都提出了一些公理作为概率论的前提，但并不完善。事实上，通过大批数学家的努力和对概率论基本概念的分析，数学家们认识到：真正严格的公理化概率论只能建立在测度论与实变函数理论之上。在这方面的先行者是法国数学家博雷尔。博雷尔是测度论的奠基人，他早在 1905 年就认识到概率论理论应建立在测度论的基础上，并首先将测度论方法引入到概率论重要问题的研究中。

伯恩斯坦（Bernstein，1880~1968）是原苏联数学家。1880 年 3 月 6 日生于敖德萨，1968 年 10 月 26 日卒于莫斯科。

伯恩斯坦 1893 年毕业于法国巴黎大学，1901 年又毕业于巴黎综合工科学校。1904 年在巴黎获数学博士学位，1907 年成为教授。1914 年在哈尔科夫又获纯粹数学博士学位。1907~1933 年在哈尔科夫大学任教，1933~1941 年在列宁格勒综合技术学院和列宁格勒大学工作，1935 年以后在原苏联科学院数学研究所工作。1925 年当选为乌克兰科学院院士，1929 年当选为原苏联科学院院士。他还是巴黎科学院的外国院士。伯恩斯坦曾获得许多国家的荣誉称号和奖励。

伯恩斯坦对偏微分方程，函数构造论和多项式逼近理论，概率论都作出了贡献。伯恩斯坦对变分法、泛函分析等也有贡献。

在数学中以他的姓氏命名的有：伯恩斯坦定理、伯恩斯坦多项式、伯恩斯坦不等式、伯恩斯坦插值法、伯恩斯坦拟解析类、伯恩斯坦求和法、伯恩斯坦 - 科尔莫哥洛夫估计、伯恩斯坦 - 佐滕多项式、伯恩斯坦极小子流形问题等，而其中以他的姓氏命名的定理有多种。伯恩斯坦的主要论著都被收入 1952~1964 年出版的《伯恩斯坦文集》1~4 卷中

在概率论方面，他最早（1917 年）提出了一些公理来作为概率论的前提，促进了概率

论公理化的建立。他与莱维共同开创了相关随机变量之和依法则收敛问题的研究。1917 年他们得到了相当于独立随机变量之和的中心极限定理，其特点是把独立性换为渐近独立性。从 1922 起，他又着手研究一些应用的实例，诸如马尔可夫单链成果的推广等。他与莱维在研究一维布朗扩散运动时，曾尝试用概率论方式研究所谓随机微分方程，并可将它推广到多维扩散过程的研究。

1917 年伯恩斯坦发表了一篇论文，题目是“论概率论的公理化基础”，随后的几年里他仍致力于研究概率论公理化。1927 年他的《概率论》第一版出版，最后一个版本，即第四版本出现于 1946 年。许多年来这本书不仅是数学家、物理学家的参考书，也是许多其他学科学生的教材。伯恩斯坦在书中给出来一个详细的概率论公理体系。他假定我们在自然科学中的推理是基于以往的经验，只要给定的条件集合 α 实现，属于已知类 A 的一个事件必然发生，这和其他因素无关。然而，一般而言，一个事件不可能绝对出现。人们不能完全确切地预言真实现象的行为。只有当条件集合 α 不太大，而且易于观察时，把 α 和 A 联系起来的规律才有世纪意义。如果这个条件不成立，事件 A 就叫做随机事件。然后他试着引进一个简单点的条件集合 β 来代替 α，它（至少在理论上）可以重复实现无限多次，当 β 存在时，给定试验中事件 A 以一个明确的概率发生，而且这个概率可以用数值表示。如果也定义了事件 B 的概率，那么下面三个关系必有一个成立

$$P(A)=P(B);P(A)>(B);P(A)<P(B)$$

然后，伯恩斯坦引进了三个公理：

(1)概率的可比较性公理。

(2)不相容事件公理。

(3)事件组合公理。

前两个公理考虑了条件集合 β 固定的情况，第三个公理把条件 α 下 A 的概率与不同的条件集合 β 下同一事件的概率联系起来。

伯恩斯坦就在这三个公理的基础之上构造了概率论的整个大厦。正像柯尔莫戈洛夫所指出了：“第一个系统的概率论公理化体系是伯恩斯坦给出的，他建立的基础是，根据随机事件的概率对事件做定性比较的思想。在定性比较这一思想中概率的数值似乎是推导而来的，而不是一个基本概念。”

米泽斯（Mises，1883~1953)是奥地利数学家、空气动力学家。1883 年 4 月 19 日生于奥地利伦贝格，1953 年 7 月 14 日卒于美国波士顿。

米泽斯 1907 年在维也纳获博士学位。1909~1918 年任斯特拉斯堡大学应用数学教授，1920 年任柏林大学应用数学教授和应用数学所所长，1921~1933 年同时是《应用数学与力学杂志》的创始人与编辑。纳粹上台后，他 1933 年离开德国到土耳其伊斯坦布尔大学任教，1939 年到美国哈佛大学任教，1944 年任该校戈登麦凯空气动力学和应用数学教授。

米泽斯是概率的频率理论学派的代表人物。他继承 19 世纪频率理论的先驱者泊松和维恩等人的思想，把一事件的概率定义为该事件在独立重复随机试验中出现的频率的极限，并把此极限的存在性作为他的第一条公理。他的第二条公理是，对随机选取的子试验序列，事件出现的频率的极限也存在并且极限值相等。这种频率法的理论依据是强大数律，它具有较强的直观性，易为实际工作者和物理学家所接受，但他给出的定义排斥了在概率论中十分重要的诸如某个事件在一无限重复的试验序列中无穷多次发生的概率，虽经他本人及后人多方修饰，仍不尽如人意，仍不断有人在继续讨论研究，真正严格的公理化概率

论只有在测度论与实变函数的基础上才可能建立。他建立的频率的极限理论反映在其编著的《概率，统计和真理》一书中。

米泽斯早期的工作集中于空气动力学，尤其是边界层流理论和机翼设计理论。1915 年，由他设计、奥地利军队制造了一架 600 马力的军用飞机，而且他作为驾驶员在第一次世界大战中服役。他关于飞行理论的著作曾一再增订出版。另外，他对弹性、塑性、湍流理论、数值分析等学科也有贡献。

概率概念是公理化概率论的基础。很长时间以来人们已经认识到概率的古典定义的缺点。同样，概率的主观解释也有严重的错误。上面已经提到人们对此有着不同的理解。当时出现了许多对这些缺点的评论，其中最著名的是米泽斯的工作。他的主要工作是概率论的频率定义和统计定义的公理化，在 1928 年的《概率，统计和真理》一书中建立了频率的极限理论。他和他的频率学派明确强调，概率概念只有在大量现象存在时才有意义米泽斯的频率理论中最根本的概念是“集体”概念。集体是相似的事件或过程组成的无限序列 K，每个事件确定一个给定的有限维空间 R 的某个点。他把一个事件的概率定义为该事件在独立重复随机试验中出现的频率的极限，并把此极限的存在性作为他的第一条公理。他的第二条公理是，对随机选取的子试验序列，事件出现的频率的极限也存在并且极限值相等。严格来说，这第二条公理没有明确的数学含义。因此，这种所谓公理化在数学上是不可取的。虽然频率定义在直观上易于理解，易为实际工作者和物理学家所接受，便于在实际工作中应用。但是，像某个事件在一独立重复试验序列中出现无穷多次这一事件的概率，米泽斯理论是无法定义的。因此，他们的公理理论是不尽人意的。

博雷尔（Borel，1871~1956）是法国数学家。1871 年 1 月 7 日生于圣阿弗里克，1956 年 2 月 3 日卒于巴黎。

博雷尔 1893 年毕业于巴黎高等师范学校，1894 年获博士学位。自 1893 年毕业后，先后在大学（里尔大学、巴黎高等师范学校、巴黎大学理学院等）任教近 50 年之久。他除任教外，还担任过巴黎高等师范学校和安利普安卡尔大学的校长。1921 年被选为法国科学院院士。1928 年组建庞加莱研究所，并任所长直至去世。他还是前苏联科学院和其他科学院的外籍院士。

博雷尔是 20 世纪第一流的数学家。他对数学分析、函数论、数论、代数、几何、数学物理、概率论等诸多分支都有杰出的贡献，他是一位多产的数学家。法国数学家弗雷歇格（Frechet）曾说：“仅仅为了归纳，简述博雷尔的作品就需要数卷篇幅”。在他不下 300 种作品中，有 30 余本著作多次再版，不少译成外文，他还多次获法国科学院奖。

博雷尔发展了现代数学分析的不同方向。1895 年以来他深入地研究了发散级数，可和性级数理论的系统发展就是由他开始的。他引进了绝对可和性的概念，并证明了绝对可和的发散级数可以完全像收敛级数那样进行运算。换句话说，这种级数代表一个函数，并且可以像函数一样进行运算。例如，两个绝对可和级数的和、差、积仍是绝对可和的，并且分别是每个级数所代表的函数的和、差、积，类似地对绝对可和级数的微商也成立。他曾借助于级数来研究任意函数。他写的《发散级数论》（1899）年获得法国科学院大奖。博雷尔清楚地认识到从一个区间的所有开覆盖中能够选出有限个覆盖的重要性。他完善了海涅（Heine）提出的覆盖定理，即现在的所谓“海涅-博雷尔定理”或“有限覆盖定理”，此定理和戴德金的“分割”法则、区间套定理、波尔察诺-魏尔斯特拉斯聚点存在定理是等价的。

博雷尔1898年改进了容度的概念，提出了测度的概念，从而发展了测度理论。博雷尔还是最先注意到康托尔思想重要性的一位数学家，并首先把康托尔的思想用于函数论。他的主要著作发表在《函数论专集》中，这部专著对后来函数论的研究产生过重要的影响。博雷尔发展了解析函数。例如，他引进了单演函数，这种函数是复平面的一般点集上的可微函数。博雷尔对概率论也有深入的研究，他把概率论同测度论相结合，引进了可数事件集的概率，填补了古典有限概率和几何概率之间的空白。他对对策论也极有建树：他首次定义了策略的应对，考虑了最优策略、混合策略、均衡策略和无限策略，并应用于战争及经济建设。

博雷尔不仅是一位杰出的数学家，而且还是一位著名的社会活动家。他曾任法国国民议会议员（1924~1936年）、海军部长（1925~1940年）。还当过市长。第二次世界大战期间参加抵抗运动，并荣获抵抗运动奖章（1954年）。他曾被授予大十字军功章（1950年）和国家科学研究中心颁发的第一枚金质奖章。法国数学家蒙泰尔（Montel）曾说："博雷尔的思想将会长久地继续在研究中发挥影响，就像远处的星光散布到广阔的空间"。

博雷尔1920年曾到中国进行过学术访问，并逗留了五个月。

在数学中以他的姓氏命名的有：博雷尔函数、博雷尔测度、博雷尔变模、博雷尔集、博雷尔强大数律、博雷尔同构、博雷尔定理等。

从20世纪开始，概率论的研究类型在很大程度上是由集合论和函数论的思想所决定的。通过对概率论基本概念——事件与概率的仔细分析，可以发现事件的运算与集合的运算完全类似，概率与测度有相同的性质，这成为建立概率论逻辑基础的正确道路。

博雷尔的工作引起了数学家们沿着这一思路的一系列探索，尤其是前苏联数学家柯尔莫戈洛夫的工作最为突出。其实，现在通行的概率论教科书中关于概率的公理化定义就是柯尔莫戈洛夫于1933年提出的。

3.3 柯尔莫戈洛夫的《概率论基础》研究

1933年，柯尔莫戈洛夫的《概率论基础》出版，这是概率论发展史上的具有"里程碑"式的著作。

柯尔莫戈洛夫(A.H.Konmoropos，1903~1987)，苏联数学家。1903年4月25日生于俄罗斯坦波夫市。他原籍是雅罗斯拉夫州图诺申诺，父亲是当地农庄的农艺师。母亲玛利亚·雅可夫列夫娜在回庄园途中临产，因汽车受阻而停留在坦波夫，不幸在那里难产死去。从此，由玛利亚的妹妹薇拉·雅可夫列夫娜承担抚养和教育小柯尔莫戈洛夫的重任。小柯尔莫戈洛夫在图诺申诺度过了幼年，6岁那年随姨妈迁居莫斯科。

柯尔莫戈洛夫的健康成长，得感谢这位好姨妈。薇拉是位有教养、有崇高社会理想的自立的女性，她的思想对外甥很有影响，在强烈的求知欲和独创精神方面尤为显著。少年的柯尔莫戈洛夫兴趣广泛，对历史、生物、人文等方面的书都很爱研读。1910年，柯尔莫戈洛夫进莫斯科普列曼私立古典中的预备班学习。这是一所由文化素养较高的知识分子办的学校，柯尔莫戈洛夫在这里受到良好的教育。

柯尔莫戈洛夫少年时已显露出很强的自主性，能妥善地支配时间抓紧学习。12至14岁时，他按《百科全书》的内容有系统地自学高等数学，读了有关社会制度、立宪议会知识方面的书。1917年10月革命风暴掀起，14岁的柯尔莫戈洛夫为社会变革中出现的一些新事物所吸引，参加了立宪议会选举。后来，他回忆起这段往事，认为对一名14岁的孩

子来说，参加这些活动对了解社会，培养对社会的责任感颇有影响。

在进入莫斯科大学之前，柯尔莫戈洛夫已经工作，他在铁路上当列车员、锅炉工、车厢图书管理员等。他随列车遍游了俄罗斯。这一职业使他能接触各阶层人民群众，饱赏大自然风光景色，这对后来柯尔莫戈洛夫在科学活动中表现出的一种视野开阔的品格，不可能没有影响。

开始，柯尔莫戈洛夫并没有把数学作为自己的志向，他对历史感兴趣。进莫斯科大学后，他很有兴致地参加著名历史学家巴赫鲁申教授主持的讨论班，并依据 15~16 世纪留下的土地财产簿对诺夫戈德地区的土地关系进行研究，又对古代上游某地区的移民路线作出推测。后来，专家组织考察，证实了柯尔莫戈洛夫的推测是正确的。这些研究成果，足以说明柯尔莫戈洛夫对人文科学的研究能力。同时，对工程建设的向往，使他进莫斯科大学时，又成为化工学院冶金专业的学生，并自信能成为一名冶金师。

而这段时间他在数学领域也出了成果。1922 年，在前辈学者、莫斯科学派的领导成员鲁金的影响下，他完成了第一篇论文《关于傅立叶系数的阶》。在这篇文章里，柯尔莫戈洛夫构造了几乎处处发散的傅立叶级数的例子，接着又构造了每一点都发散的傅立叶级数的例子。这些例子使专家受到震动，对深入研究函数论有很大作用。

这一成果，促使柯尔莫戈洛夫最终决定转向数学，并成为鲁金的高材生。1925 年，他以优异的成绩毕业，此后继续在鲁金主持的研究班里学习和开展学术研究，从此也开始了他大学教师的生涯。

自柯尔莫戈洛夫 19 岁发表第一篇论文到 84 岁去世为止的年中，他研究涉足的领域包括：函数论、概率论、数理逻辑、湍流力学、遍历理论、动力系统理论、信息论、自动控制理论、数学史等。他的贡献归纳起来大致有以下几个方面。

(1)概率论：早在 1924 年，柯尔莫戈洛夫就开始以测度和实变函数论为基础，把概率论的发展推进到一个崭新的阶段，为随机过程的理论提供了必要的基础。1928 年，他证明了大数定律的必要和充分条件，同时还证明了在项数上加上极宽条件时有关独立随机变量的重对数法则，此外，他推广了古典的切比雪夫不等式，提出著名的柯尔莫戈洛夫不等式。

1931 年，柯尔莫戈洛夫发表了《概率论的解析方法》，为现代马尔可夫随机过程和揭示概率论与常微分方程及二阶偏微分方程的内在联系奠定了基础。这些理论对力学、生物学、化学和工程技术都有价值，迅速成为现代自然科学的有力工具。

1933 年，柯尔莫戈洛夫发表了《概率论基础》一书，建立了概率公理化体系，给出了无穷可分律的分布表达式，提出了经验分布和实际分布的最大偏差的极限法则，即柯尔莫戈洛夫准则，创立了具有可数集状态的马尔可夫理论。

(2)湍流力学：柯尔莫戈洛夫重视数学的应用，常常把抽象的数学理论与自然科学实验融为一体。30 年代末，柯尔莫戈洛夫对气流、液流运动的一些规律问题进行研究，用统计力学和建立随机函数的方法对这些力学现象作严格的数学描述，创立了统计流体力学。他在这方面的主要著作发表于 1941 年，文章引用函数空间的测度，对具有很大自由度且极端复杂的非线性系统演变过程的木质开展研究，得出这一演变规律的数量关系，提出著名的三分之二法则，即在任意高度的湍流运动中，距离的两个质点运动的均方差正比于 rsl.

(3)数理逻辑：战后，柯尔莫戈洛夫研究了数理逻辑基础及其在几何学，概率论，信息论中的应用，他早年和晚年都钻研过数理逻辑，包括算法论和数学基础。早在 1925 年，他在《数学文集》上发表了关于排中律的文章。文章提出潜入运算的概念与方法，现在被称为柯尔莫戈洛夫运算。1932 年，他发表第二篇关于直观逻辑的文章，提出把直观逻辑解

释为结构逻辑的可能性。

1952年，柯尔莫戈洛夫提出物质构造的最普遍的定义和算法的最一般定义。1954年的文章形成读数法理论的最初概念。1972年，他在莫斯科大学数学系首先开设数理逻辑必修课。

(4)动力系统理论：50年代初，柯尔莫戈洛夫研究了动力学系统理论。这一课题是牛顿、拉普拉斯、庞加莱等学者研究的延续与发展，他的文章《动力学的攀本问题》涉及哈密顿系统的一般理论以及关于天体运行的三体和多体问题。柯尔莫戈洛夫将信息论的思想用于动力系统，这体现在他与他的学生阿诺尔德合写的关于摄动哈密顿系统的著作里，这本著作获得19年列宁奖金。

在《过渡的动力系统与勒贝格同构空间的新度量的变量》一文中，柯尔莫戈洛夫引进熵的特征值的概念，文章对动力系统的遍历定理的发展起着特殊的作用，由此开创了一系列新课题的研究。文章提出的拟正则概念，后来被学术界为纪念他的这一新观点而命名为"K—系统"。柯尔莫戈洛夫还把遍历理论的思想用于研究湍流型的力学现象并取得成功。

(5)数学史：在柯尔莫戈洛夫的著作中，数学史占显著地位。他提出数学发展的时期划分，对每一时期数学发展的动力及科学家的贡献给以评价，对各时期所积累的经验也给以概括。他还专门考察了概率论这一分支的发展分期，对俄罗斯科学家如切比雪夫、马尔可夫、李雅普诺夫等对概率论的成熟及发展所作的特殊贡献给以颂扬。此外，对罗巴切夫斯基创立非欧几何之后，在数学思维方法上所引起的变化作了深刻的分析。柯尔莫戈洛夫还写文章评价了德国数学大师康托尔、希尔伯特的思想和成就。他还担任过《19世纪数学史》丛书的编辑。总的来说，柯尔莫戈洛夫科学研究的特点是，既有工作重心，又能四面出击，他常把几个领域的成果相互渗透、左右蔓延，向国民经济和技术部门扩大战果。他既是理论家，又是实践家，他曾亲自参加海洋考察队和军事上关于炮火程序控制方面的研究，在研究领域的广泛性与交叉性方面极为突出。

柯尔莫戈洛夫一贯重视教育，他既是科学家，又是教育家。他亲自讲课，编写大纲、教材，主持讨论班等，使一代又一代人在他精心培育下成长。

1922年，19岁的柯尔莫戈洛夫就在俄罗斯联邦人民教育委员会实验中学兼课，开始了教师生涯。在这所中学他一直工作到1925年大学毕业时止。他热爱教师工作，具有良好的教师素质。1931年，他成为莫斯科大学教授，1933至1939年还担任大学的数学研究所所长。他特别注意培养青年，善于在大学生和研究生中选拔人才，让他们参加由他主持的讨论班。由于柯尔莫戈洛夫的基础扎实，知识渊博，对所研究的课题具有深刻的洞察力，每次讨论都进行得很热烈，参加者受益匪浅。在柯尔莫戈洛夫的学生中有7名国家科学院院士，5名通讯院士，70名加盟共和国科学院院士和一大批享有盛誉的学者。

柯尔莫戈洛夫爱好文学，喜欢欣赏绘画、雕刻、建筑艺术。此外，他还是一名卓越的滑雪运动员和游泳能手。课余时，他常参加徒步旅行。在假日喜欢组织3至5名研究生或大学生到莫斯科郊外旅游。郊游中除了讨论数学问题之外，还谈论沿途的建筑、雕塑、名胜古迹和文化生活中的突出事件，与青年人共进午餐。通过对人类智慧和文明的赞颂，增强了他们探求科学、创造美好未来的信念。参加者心境开阔，许多青年成才之后，对昔日的郊游活动总是记忆犹新，难以忘怀。柯尔莫戈洛夫还是教学活动的组织者，他的工作中心一直在莫斯科大学。1938至1966年主持由他创建的概率论教研室的工作，1966年至1976年主持由他创建的统计方法试验室的工作，1976年至逝世前担任数理统计教研室主任和数理逻辑教研室主任。

近 20 多年来，柯尔莫戈洛夫关心中学教育。在莫斯科第 18 寄宿中学(即柯尔莫戈洛夫中学)的筹建和创办时期，他倾注了心血，亲自到校讲课，上习题课。在他指导下编写的中学教材《代数与初等分析》成为苏联通用教科书。柯尔莫戈洛夫还饶有兴趣地为中学生开设音乐、绘画和文学讲座，体现了老教育家既重视开发学生的智力，又重视美学的远见卓识。这所中学为莫斯科大学输送了大批本科生和研究生。学生曾多次获得全苏和国际奥林匹克竞赛奖。

几十年来，柯尔莫戈洛夫的科学活动在国内外都得到很高的评价。他 7 次获得列宁勋章，被授予苏联社会主义劳动英雄称号，是列宁奖金和国家奖金的获得者。1963 年，他荣获国际巴尔奖，1980 年获得国际数学界的崇高奖赏—沃尔夫奖。

柯尔莫戈洛夫曾被选为 20 多个国外学术团体的院士，会员或荣誉会员，其中有荷兰皇家科学院、伦敦皇家科学院、美国国家科学院、巴黎科学院、罗马尼亚科学院。能得到这么多的荣誉，在国际数学界是首屈一指的。在柯尔莫戈洛夫 70 周年、75 周年和 80 周年诞辰时，苏联科学界都组织过纪念活动。苏联数学学会和《数学科学成就》等杂志发表对他的颂文。在 1978 年庆祝他 75 周岁生日的集会上，主持者所致的颂词中说："您不但以自己的真知灼见丰富了上述这些数学领域，而且发展了数学的应用。您一贯重视数学的应用及其与其他的联系，您在学术上的兴趣是非常广泛的。您有当教师的天才，总是吸引着有才华的青年们。您建立了全世界闻名的科学学派，这个学派的代表人物正在数学各个领域和其他自然科学领域内卓有成效地工作着"。

1987 年 10 月 20 日，柯尔莫戈洛夫与世长辞。他所走过的宏伟壮丽的教学和科研历程，永远鼓舞着无数后来者去开创更加美好的未来。

柯尔莫戈洛夫从 20 世纪 20 年代中期开始沿着测度论途径研究整个概率论理论的严格表述，其结果发表在他在 1933 年以德文出版的经典著作《概率论基础》。柯尔莫戈洛夫是莫斯科函数论学派领袖鲁金的学生，对实变函数理论掌握的可谓炉火纯青。他在这部著作中建立起集合测度与事件概率的类比、积分与数学期望的类比、函数正交性与随机变量独立性的类比等。这种广泛的类比终于使概率论具有了演绎数学的特征。柯尔莫戈洛夫提出了 6 条公理，整个概率论大厦可以从这 6 条公理出发建立起来。柯尔莫戈洛夫公理化概率定义的实质是把概率定义为随机试验样本空间上的集合函数。柯尔莫戈洛夫公理化体系逐渐得到了数学家们的普遍承认。由于公理化，概率论成为一门严格的演绎数学，具有与其他数学分支分庭抗礼的地位，并通过集合论与其他数学分支密切地联系着，开创了现代概率论新时期。

1.概率概念的完善

概率是描述事件发生(出现)可能性大小的数量指标，它是逐步形成和完善起来的。最初人们讨论的是古典概型(随机)试验中事件发生的概率。所谓古典概型试验是指样本空间中的样本点的个数是有限的且每个样本点(组成的事件)发生的可能性是相同的，简称为有限性与等可能性。例如，掷一颗均匀骰子的试验与从一个装有 n 个相同(编了号)的球中随机摸一个球的试验都是古典概型试验。对于古典概型试验，人们给出概率的如下定义。

定义：设试验 E 是古典概型的，其样本空间 Ω 由 n 个样本点组成，其一事件 A 由 r 个样本点组成，则定义 A (发生)的概率为 $\frac{r}{n}$，记为 $P(A)$，即

$$P(A)=\frac{A\text{中样本点数}}{\Omega\text{中样本点数}}=\frac{r}{n}, \tag{3-1}$$

并称这样定义的概率为古典概率，称概率的这样的定义为古典定义。

古典概率有如下 3 个性质：

(1)对任意事件 A，有 $0 \le P(A) \le 1$。

(2) $P(\Omega)=1$。

(3)设 $A_1, A_2, \cdots, A_m$ 为两两互斥的 m 个事件，则

$$P(\bigcup_{i=1}^{m} A_i)=\sum_{i=1}^{m} P(A_i)$$

性质(1)、(2)、(3)分别称为概率的有界性规范性与有限可加性。

古典概率的定义要求试验满足有限性与等可能性，这使得它在实际应用中受到了很大的限制。例如，对于旋转均匀的陀螺的试验：在一个均匀的陀螺圆周上均匀地刻上区间$[0,3)$内诸数字，旋转陀螺，当它停下时，其圆周上与桌面接触处的刻度位于某区间$[a,b)$ $(\subset[0,3))$内的概率有多大?对于这样的试验，古典概率的定义就不适用。因为此试验的样本点不是有限的，而是区间$[0,3)$中的每个点，它有无穷不可数多个。为了克服古典概率定义的局限性，人们又引入概率的如下定义。

定义：设试验 E 的样本空间为某可度量的区域Ω，且Ω中任一区域出现的可能性大小与该区域的几何度量成正比，而与该区域的位置与形状无关，则称 E 为几何概型的试验，且定义 E 的事件 A 的概率为

$$P(A)=\frac{A\text{的几何度量}}{\Omega\text{的几何度量}}, \tag{3-2}$$

其中，如果Ω是一维的、二维的、三维的，则Ω的几何度量分别为长度、面积、体积。并称这样定义的概率为几何概率，而称概率的这样的定义为几何定义。

几何概率除了具有古典概率的三个性质外，它还具有以下的可列可加性(或完全可加性)：

(4)设 $A_1, A_2, A_3, \cdots$ 为两两互斥的无穷多个事件，则

$$P(\bigcup_{i=1}^{\infty} A_i)=\sum_{i=1}^{\infty} P(A_i)$$

概率的几何定义虽然去掉了有限性的限制，但是它仍然要试验满足等可能性，这在实际问题中仍有很大的局限性。例如，掷一枚不均匀的硬币的试验就具有等可能性，这样上述两个定义对这个非常简单的试验都不适用。同时我们还注意到上述两个定义中的等可能性严格地说都是近似的，而不是真正的等可能。为此，人们在频率的基础上又引进了概率的统计定义。

通过长期的实践，人们逐步发现，当重复实验的次数很多时，事件出现的频率都具有稳定性。即对于某个固定的事件，当重复试验次数增加时，该事件出现的频率总在 0 与 1 之间某个数字 p 附近摆动，且越来越接近 p。例如，掷一枚均匀硬币的试验，历史上曾经有很多数学家做过，表 3-1 是几位数学家做此试验的结果。

表 3-1

试验者	试验次数	正面出现次数	正面出现频率	试验者
摩根	2 048	1 061	0.5181	摩根
蒲丰	4 040	2 048	0.5069	蒲丰

续表

皮尔逊	12 000	6 019	0.5016
皮尔逊	24 000	12 012	0.5005
维尼	30 000	14 994	0.4998

由表 3-1 可以看到，当试验次数越来越多时，正面出现的频率也越来越靠近 0.5。由此，人们又引入概率的统计定义。

定义：设 A 为试验 E 的一个事件，如果随着重复试验次数的增加，A 出现的频率在 0 与 1 之间某个数 p 附近摆动，则定义 A 的概率为 p，记为 $P(A)$，即

$$P(A)=p,$$

称这样定义的概率为统计概率，称概率的这样的定义为统计定义。

统计概率也有古典概率的三个性质，即有界性、规范性、有限可加性。

概率的统计定义对试验不作任何要求，它适合所有试验，也比较直观。但是在数学上很不严格，因为其依据是重复试验次数很多时频率呈现出的稳定性。何为“很多”?1 万次相对于 1 000 次来说是很多了，但是相对于 10 万次来说它又是很少了。试验次数究竟要多到怎样的程度才能算“很多”，定义中没有说明；又如定义中的“摆动”又如何理解，也没有数学说明，再如定义中的“p”又如何确定？不同的人可能会确定不同的“值”。这样，一个事件将有多个概率。例如，在表 3-1 中，正面出现的频率显然在 0.5 附近摆动，因此可以认为正面出现的概率为 0.5。但是由于硬币不会绝对均匀的，也可以认为正面出现的概率为 0.50001 或 0.4999。因此，概率的上述三个定义都有缺陷，与其说它们是定义，不如说它们仅仅是对不同的情况给出概率的三种计算方法。所以我们有必要给出概率的一个严密的对各种情况都适用的定义，以使得概率论这座大厦有牢固的基础。

20 世纪 30 年代初，米泽斯给出样本空间的概念，使得有可能把概率的严密的数学理论建立在测度论上。20 世纪 30 年代中期柯尔莫戈洛夫以上述三个定义的性质为背景给出概率的严密的公理化定义。

定义：设 $(\Omega,\Re)$ 为一个可测空间，P 为定义于 $\Re$ 上的实值集合函数，如果 P 满足下列三个条件：

(1)对每个 $A\in\Re$，有 $P(A)\geq 0$。

(2) $P(\Omega)=1$。

(3)如果 $A_i\in\Re, i=1,2,3,\cdots$，且当 $i\neq j$ 时，$A_iA_j=\Phi$，则 $P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$。

那么，就 P 称为概率测度，简称概率。

一般地把 $\Omega,\Re,P$ 写在一起成 $(\Omega,\Re,P)$，并称 $(\Omega,\Re,P)$ 为概率空间，以后总用 Ω 表示样本空间，用 $\Re$ 表示 Ω 中的固定的事件域，用 P 表示相应于 Ω 与 $\Re$ 的概率。此定义的三个条件称为三个公理。这三个公理分别称为概率的非负性、规范性与完全可加性(或可列可加性)。

这样定义的概率 P 有如下性质：

(1)(不可能事件 Φ 的概率为零) $P(\Phi)=0$。

(2)(有限可加性) 设 $A_i\in\Re, i=1,2,3,\cdots,n$，且当 $i\neq j$ 时，$A_iA_j=\Phi$，则

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)$$

(3)(对立事件概率公式)设 $A\in\Re$，则

$$P(\overline{A})=1-P(A)$$

(4)(正常差概率公式)设 $A,B\in\Re$，且 $A\subseteq B$，则

$$P(B-A)=P(B)-P(A)$$

(5)(单调性) 设 $A,B\in\Re$，且 $A\subseteq B$，则

$$P(A)\le P(B)$$

(6)(有界性) 设 $A\in\Re$，则

$$0\le P(A)\le 1$$

(7)(加法公式) 设 $A_i\in\Re, i=1,2,\cdots,n$，则

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j}^{n}P(A_iA_j)\ \sum_{1\le i<j<k}^{n}P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)$$

特别地，

$$P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1A_2),$$

$$P(A_1\cup A_2\cup A_3)=\sum_{i=1}^{3}P(A_i)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)$$

(8)(半有限可加性) 设 $A_i\in\Re, i=1,2,\cdots,n$，则

$$P(\bigcup_{i=1}^{n}A_i)\le\sum_{i=1}^{n}P(A_i)$$

(9)(半完全可加性) 设 $A_i\in\Re, i=1,2,\cdots$，则

$$P(\bigcup_{i=1}^{\infty}A_i)\le\sum_{i=1}^{\infty}P(A_i)$$

以上性质都可以由三个公理推导而来。

概率的公理化定义中没有要求定义于 $\Re$ 上的实值集合函数 P 满足有界性与有限可加性，为什么?这是因为有界性与有限可加性可以由三个公理推导出来，而且，一个概念的定义(自然)要求所满足的条件越少越好，这样才便于应用。设想，如果一个定义要求满足10个条件，则每次应用前都要逐一验证这10个条件是否满足，这将是很麻烦的事情。其次，概率的公理化定义是严密的数学定义，且对试验不作任何要求，我们很自然地会问，前述的三个定义是否可以不要了?不可以。这是因为公理化定义虽然在数学上很严密，但是它没有给出事件概率的计算方法。要计算一个具体的事件的概率，还得根据不同的情况，利用上述三个定义之一来计算。

2.其他公理体系

概率的公理化定义不是唯一的，它有很多等价定义。介绍概率论历史发展的文章，或者概率论的教科书一般只提到了柯尔莫戈洛夫建立的公理体系。这和柯尔莫戈洛夫的工作的影响不无关系，但同时容易使人造成一种误会，以为概率论只有一种公理化结构。其实不然，根据人们对概率概率的不同理解、体会，可以构造出不同的公理化理论。

例如，Renyi 公理系统：设 Ω 表示整个基本事件空间的集合，γ 表示 Ω 子集的 σ 代数，γ 的要素 A，B，…称为事件，设表示一个非空集合，并且有 $\subseteq\gamma$。在 $A\in\gamma,B\in\gamma$ 下定义 $P\{A\mid B\}$。

Renyi 公理系统表述如下：

(1)对于任意给定事件 A,B，有

$$P\{A|B\}\ge 0\text{，且 }P\{B|B\}=1.$$

(2)对于不相容事件 $A_1,A_2,\cdots$ (不会发生在同一试验中的事件)，和某事件 B，存在

$$P\{\bigcup_{i=1}^{\infty}A_i|B\}=\sum_{i=1}^{\infty}P\{A_i|B\}$$

(3)对于每个事件集合 (A,B,C)，$B\subseteq C$，$P\{B|C\}>0$，存在

$$P\{A|B\}=\frac{P\{A\cap B|C\}}{P\{B|C\}}$$

注意：如果 $P\{A\mid B\}$ 不依赖于 B，则称 A 和 B 是相互独立的。

Renyi 的条件概率公理系统，从主观角度来解释概率函数 $P(A)$，也就是说，将 $P(A)$ 理解为对事件 A 的置信度。Renyi 的系统的理论基础是实用主义：现实生活中大多数判断和决策都是受以往信息和经验制约的，而且只需引进一个公理系统来处理不可测事件。

3.4 概率论公理化以后

在概率论公理化基础上，现代概率论取得了一系列理论上的重大突破。一是随机过程的研究由马尔可夫过程推进到对一类特殊的马尔可夫过程——布朗运动的研究；二是对具有重要意义的随机过程“鞅”的研究；三是由日本数学家伊藤清于 1942 年引入的随机积分与随机微分方程的研究。

由于科学技术中许多实际问题的推动以及概率论逻辑基础的建立，概率论从 20 世纪 30 年代以来得到了迅速的发展。

目前其主要研究内容大致可分为极限理论，独立增量过程，马尔可夫过程，平稳过程和时间序列，鞅和随机微分方程，点过程等。此外，包括组合概率（用组合数学方法解决只涉及有限个基本事件的概率问题）、几何概率等在内的一些属于古典范畴的问题，至今仍有人在继续研究，并有新的发展。

极限理论是研究与随机变量序列或随机过程序列的收敛性有关的问题的理论。20 世纪 30 年代以后，有关随机变量序列的极限理论（主要是中心极限定理）的研究，是将独立序列情形的结果推广到鞅差序列和更一般的弱相依序列等情形，以及研究收敛速度问题。近年来，由于统计力学的需要，人们开始研究强相依随机变量序列的非中心极限定理。自 1951 年 M.唐斯克提出不变原理（见随机过程的极限定理）后，有关随机过程序列的弱收敛的研究成了极限理论的一个中心课题。普罗霍洛夫及 A.B.斯科罗霍德在这方面作出了最主要的贡献。1964 年 V.斯特拉森的工作出现后，引起了有关随机过程序列的强收敛的研究，这就是强不变原理。近年来，鞅论方法已渗透到这一领域，使许多经典结果的证明得到简化和统一处理，并且还导致一些新的结果。

人们最早知道的独立增量过程是在物理现象中观察到的布朗运动和泊松过程，一般的独立增量过程的研究，归功于莱维，它在 20 世纪 40 年代已臻成熟。在这些研究中，包含了许多重要的方法和概念，概率论的许多近代研究课题都直接或间接地受其启发与影响。

在实际中遇到的很多随机现象有如下的共同特性：它的未来的演变，在已知它目前状态的条件下与以往的状况无关。描述这种随时间推进的随机现象的演变模型就是马尔可夫过程。

20 世纪 50 年代以前，研究马尔可夫过程的主要工具是微分方程和半群理论（即分析方法）；1936 年前后就开始探讨马尔可夫过程的轨道性质，直到把微分方程和半群理论的分析方法同研究轨道性质的概率方法结合运用，才使这方面的研究工作进一步深化，并形成了对轨道分析必不可少的强马尔可夫性概念。1942 年，伊藤清用他创立的随机积分和随机微分方程理论来研究一类特殊而重要的马尔可夫过程——扩散过程，开辟了研究马尔可夫过程的又一重要途径。近年来，鞅论方法也已渗透到马尔可夫过程的研究中，它与随机微分方程结合在一起，已成为目前处理多维扩散过程的工具。此外，马尔可夫过程与分析学中的位势论有密切的联系。对马尔可夫过程的研究，推动了位势理论的发展，并为研究偏微分方程提供了概率论的方法。最近十多年发展起来的吉布斯随机场和无穷粒子随机系统，是由于统计物理的需要而提出的。

许多自然的和生产过程中的随机现象表现出某种平稳性。一种平稳性是过程在任意一些时刻上的联合概率分布随时间推移不变，这种平稳性称为严平稳性。严平稳过程的研究与遍历理论有密切的联系。如果上述对概率分布的要求放宽为仅对二阶相关矩的要求，即过程在任意两时刻上的协方差随时间推移不变，则称这种平稳性为宽平稳性。关于宽平稳过程的研究，辛钦、柯尔莫哥洛夫和维纳等人运用傅里叶分析和泛函分析的工具，在 40 年代已经找出了过程的相关函数及过程本身的谱分解式，并且较完满地解决了有应用意义的预测问题。许多应用问题还要求根据观测数据去建立这些数据所来自的随机过程的模型。为此产生了时间序列分析这一课题，提出了宽平稳序列的自回归滑动平均(arma)模型以及一些非线性模型。

鞅是另一类重要的随机过程。从 20 世纪 30 年代起，莱维等人就开始研究鞅序列，把它作为独立随机变量序列的部分和的推广。40 年代到 50 年代初，杜布对鞅进行了系统的研究，得到有名的鞅不等式、停止定理和收敛定理等重要结果。1962 年，P.A.迈耶解决了杜布提出的连续时间的上鞅分解为鞅及增过程之差的问题。在解决这个问题的过程中，出现了很多新鲜而深刻的概念，使鞅和随机过程一般理论的内容大大丰富起来。鞅的研究丰富了概率论的内容，并引起人们用它所提供的新方法新概念对概率论中许多经典的内容重新审议，把以往认为是复杂的东西纳入鞅论的框架而加以简化。此外，利用上鞅的分解定理，可以把伊藤清的对布朗运动的随机积分推广到对一般鞅乃至半鞅的随机积分；因而，更一般的随机微分方程的研究也随之发展。随机微分方程理论不仅可以用来研究马尔可夫过程，它还是解决滤波问题的必要工具。最近出现的流形上的随机微分方程又和微分几何及分析力学的研究发生了密切的联系。鞅论还对本学科以外的位势理论、调和分析及复变函数论等提供了有用的工具。

点过程是从所谓计数过程发展出来的，它们的特点是，可用落在不相重叠的集合上的随机点数目的联合概率分布来刻画整个过程的概率规律。最基本的计数过程是泊松过程，1943 年，C.帕尔姆将它作为最简单的输入流应用于研究电话业务问题；1955 年，辛钦又以严密的数学观点作了整理和发展。

在 60 年代以前，点过程的研究主要限于泊松过程及其推广的过程。以后，由于大量实际问题的需要以及随机测度论和现代鞅论的推动，进一步把实轴上的点过程（即计数过程）推广到一般的可分完备度量空间上，在内容和方法上都有根本性的进展。

莱维（Levy，1886~1971)是法国数学家。1886 年 9 月 15 日生于巴黎，1971 年 12 月 15

日卒于巴黎。

莱维出生于一个数学世家，祖父是大学教授，父亲是数学家。莱维在上中学时，是优秀生，得到过学校的数学奖。1902 年，年仅 16 岁时，就用比切比雪夫更简单的方法导出了素数分布密度公式，还给出了一条处处没有切线的连续曲线。1904 年考入巴黎综合工科学校学习，是阿达马的学生。毕业后到巴黎矿冶学院任教，1910 年转到赛特—埃提纳矿业学院任教 3 年，1913 年以后在综合工科学校任教，1920 年任教授。1950~1962 年先后在美国加利福尼亚、哥伦比亚、华盛顿等大学任教。1962 年返回法国，1964 年他 78 岁时当选为巴黎科学院院士。

莱维对概率论、泛函分析、拓扑学、力学都作出了贡献，尤以概率论、泛函分析为最。他重新发现并完善了特征函数理论，给出逆转公式和连续性定理（现称莱维的连续性定理）。发展了中心极限定理，提出古典中心极限定理收敛于稳定律，他提出无穷小三角序列的极限律类为无穷可分分律类（后为辛钦证明）。他提出的分布律的莱维距离、散布函数和集结函数等概念已成为研究分布律收敛的工具。他独创从样本函数角度研究随机过程，研究一般可加过程的样本函数结构，得到无穷可分分布的明显表达式。他还用随机微分方程尝试了概率方式的研究，他引进鞅的概念，证明了鞅的一些性质，并进而研究大数定律的推广。他还对布朗运动及可加过程都进行了深刻的研究，他导出了一维布朗运动关于反正弦分布定律的重要性质；在研究二维空间布朗运动曲线和其中一条弦围成的面积时，引进了由布朗运动定义的随机积分。他还引进了依赖于一个在任意有限维空间以至在可分希尔伯特空间变动的参数的布朗运动。他的工作奠定了一般极限理论和随机过程的基础。

莱维在泛函分析方面，他提出了更一般的任意一阶泛函微分方程的积分问题，不仅解决了一个泛函变元问题，还解决了对应于 $2n$ 个变元的 n 个一阶偏微分方程当 n 无限增大时的问题。他还把积分和测度的概念推广到了无限维空间，并发现了一些重要结果。“泛函分析”这个名词也是由他引进的。

莱维的主要著作有：《泛函分析教程》（1922 年）、《概率计算》（1925 年）、《随机变量的加法理论》（1937 年）、《随机过程与布朗运动》（1948 年）等。在数学中以他的姓氏命名的有：莱维不等式、莱维标准型、莱维距离、莱维过程、莱维度量、莱维连续性定理、莱维-辛钦表示、莱维-辛钦公式、莱维 - 伊藤分解定理等。

莱维成果累累，何以到 78 岁高龄才进入巴黎科学院?

1944 年，著名的分形几何创始人，法国数学家芒德布罗在一篇讨论“推测数学是否允许存在”的评论中深为不平地说道：

“历史告诉我，人类不断地产生一些数学天才，不屈服于一些常规压力，如果他们被压倒了，他们会离开数学——对所有人都是巨大的损失。”

“我的第一个证人是莱维，那时的法国数学家‘警察’一直谴责莱维没有充分地给出证明（有时是初等计算笔误）。他无法从那些数学家‘警察’手中逃脱，但他绝不改变初衷。他继续着，一直到 70 岁时，还在提供精彩绝伦和让人吃惊的直觉‘事实’——这些也许是‘不完备的’，却不断地为许多人提供了极有价值的工作。然而，当他 71 岁时(我是为他工作的初级教授)，他继续被禁止教概率论”。

辛钦（Hincen，1894~1959)是原苏联数学家。1894 年 7 月 19 日生于莫斯科附近的康德罗沃，1959 年 11 月 18 日卒于莫斯科。

辛钦 1916 年毕业于莫斯科大学并留校从事教学工作。1922~1927 年在莫斯科数学力学

研究所工作，1927 年成为教授，1932~1934 年任该所所长。1935 年获物理数学博士学位。1939 年当选为苏联科学院通讯院士，同年调到该院斯切克洛夫数学研究所工作。1944 年当选为俄罗斯教育科学院院士。他 1941 年获原苏联国家奖金，并多次获列宁勋章、劳动红旗勋章、荣誉勋章和其他奖章。

辛钦对概率论、数学分析、数论都作出了贡献。

辛钦是莫斯科概率论学派的创始人之一。他最早的概率成果是伯努利试验序列的重对数律，它导源于数论，是莫斯科概率论学派的开端，直到现在重对数律仍然是概率论重要研究课题之一，关于独立随机变量序列，他首先与柯尔莫哥洛夫讨论了随机变量级数的收敛性，他证明了：

(1)作为强大数律先声的辛钦弱大数律。

(2)随机变量的无穷小三角列的极限分布类与无穷可分分布类相同。

他还研究了分布律的算术问题和大偏差极限问题。他提出了平稳随机过程理论，这种随机过程在任何一段相同的时间间隔内的随机变化形态都相同。他提出并证明了严格平稳过程的一般遍历定理；首次给出了宽平稳过程的概念并建立了它的谱理论基础。他还研究了概率极限理论与统计力学基础的关系，并将概率论方法广泛应用于统计物理学的研究。他早在 1932 年就发表了排队论的论文。

在分析学中，辛钦早期研究成果属于函数的度量理论，他引进了渐近导数的概念，推广了当儒瓦积分，建立了辛钦积分,研究了可测函数的结构，并把函数的度量理论应用于数论和概率论中。

在数论中，辛钦的成就主要是丢番图逼近论和连分数的度量理论，建立了许多新的原理。

辛钦共发表 150 多种关于数学和数学史论著。在数学中以他的姓氏命名的有：辛钦定理、辛钦不等式、辛钦积分、辛钦条件、辛钦可积函数、辛钦转换原理、辛钦单峰性准则等，而其中以他的姓氏命名的定理有多种。他十分重视数学教育和人才的培养，潜心的编著了多本思路清晰、引人入胜、突出论题本质风格的教材和专著。其中《数学分析简明教程》、《连分数》、《费马定理》、《公用事业理论的数学方法》都已被译成中文在我国出版。

他在《数学分析简明教程》的第一版序中说：

“为了使这本教程能够尽可能地简明，我的方法完全在于选取最精简的材料，而不在叙述上压缩词句…特别是我不吝惜说一些话，来帮助读者时时刻刻都能清楚地了解到他所遵循的道路的规律。”

杜布(Doob，1910~2004)是美国数学家。1910 年 10 月 27 日生于辛辛那提。2004 年 6 月 7 日卒于伊利诺伊。杜布毕业于哈佛大学，1932 年获博士学位。他是美国国家科学院和美国科学艺术研究院院士。伊利诺伊大学教授。

杜布的主要贡献是概率论。他深入研究了随机过程理论，得出了任意的随机过程都具有可分修正,建立了随机函数理论的公理结构。他是鞅论的奠基人,虽然莱维等人早在 1935 年发表了一些孕育着鞅论的工作，1939 年维尔引进“鞅”(martingale)这个名称，但对鞅进行系统研究并使之成为随机过程论的一个重要分支的，则应归功于杜布。他还引进了半鞅的概念。在鞅论中有以他的姓氏命名的著名的杜布停止定理、杜布——迈耶上鞅分解定理等。鞅论使随机过程的研究进一步抽象化，不仅丰富了概率论的内容，而且为其他数学分支如调和分析、复变函数、位势理论等提供了有力的工具。

对马尔可夫过程，杜布关于轨道的严密处理进行了系统的研究。

他对代数函数中的聚值集的理论也作出了贡献。他还对霍普夫的个体遍历定理的特殊情形给出了证明。在数学中以他的姓氏命名的还有：杜布定理、杜布不等式、杜布收敛性等。

杜布的著作有《随机过程》(1953 年)等。

伊藤清是日本数学家，1915 年 9 月 7 日生于日本三重县。

伊藤清 1935~1938 年就读于东京大学数学系。1939~1943 年在政府统计局工作。1943~1952 年在名古屋大学任教，1943 年开始任副教授，1945 获理学博士学位。1952 年任东京大学教授。1952~1979 年在东京大学任教，但是在这 27 年中，他大约只有一半时间在东京，其余一半的时间在国外。他去过的地方有普林斯顿、斯坦福、康奈尔及丹麦。1976 年，他任日本数理解析研究所所长。1977 年当选日本学术会议会员。1979 年任日本学士院教授。他曾任日本数学会理事长。1998 年当选为美国国家科学院外籍院士。他 1978 年获日本学士院赏恩赐赏，1987 年获沃尔夫数学奖。

伊藤清在东京大学上学时，便被纯粹数学中严谨的论证与优美的结构深深地吸引了，并认识到许多数学概念都根植于力学之中。在数学与力学的天地里漫游时，通过统计力学，他最终对随机过程发生了兴趣，并参加了日本著名数学家弥尔昌吉教授主持的讨论班。在读了柯尔莫哥洛夫的《概率论的基本概念》及莱维的《随机变量的加法理论》等名著之后，打下了坚实的概率论基础。他在统计局工作期间，于 1942 年在《日本数学杂志》上发表了他的第一篇论文。在这篇论文中，他引入了刻画可微过程跳跃的泊松随机测度。后来他又在大阪大学的一份油印的杂志上发表了他的第二篇论文。在这篇论文中，他得出了决定马尔可夫过程轨道的随机微分方程的概念，它可以只借助一个可微过程的随机微分方程的微分来表示。用概率论的理论和方法研究随机微分方程，虽不是从伊藤清开始的，但他作出了系统而严密的奠基性工作。他在名古屋的前一半时间，对遍历理论、非交换群上的正定函数，以及布朗运动和调和函数之间的关系，特别是对样本轨道有兴趣。1951 年，他将维纳的齐次不规则运动稍加修改定义了多重维纳积分。1953 年，他引入了复多重维纳积分。伊藤清在名古屋的后半时期，流形理论开始吸引青年人的注意力。在这样一种气氛的感染下，他对紧流形上的扩散产生了兴趣，此扩散的生成算子是退化的椭圆形算子。他试图利用局部坐标，通过写出随机微分方程的方法来构造扩散的轨道，并相继写出了三篇论文。虽然他这三篇论文未能全部实现上述想法，但却意外地获得了随机微分的锁链法则。这一法则在理解生成算子的概率意义时是很有用的。20 世纪 50 年代后期，他与麦基恩(Mckean)合作，借助莱维的局部时概念，成功地构造了具有弹性边界的布朗运动的轨道，并合作写了一本书，名为《扩散过程及其样本轨道》（于 1965 年正式出版）。他们还合写了两篇论文：一篇是关于随机徘徊与势的；另一篇讨论了半直线上如何针对各种可能的边界条件来构造布朗运动。1956 年，伊藤清以更一般的角度定义了由可加过程导出的多重随机测度以及多重随机积分，并将其应用于种种问题。他还研究过多变数情况的广义随机过程。20 世纪 60 年代初，他获得了随机平移的概念。20 世纪 70 年代，他借助鞅积分搞清了他的积分与斯特拉托维奇积分之间的联系，并且阐明了某些数学上的应用；他还确定了马尔可夫过程在常返点的所有可能行为，从而推广了他与麦基恩共同得到的结果。在遍历理论中，他把非奇异变换的不变测度问题的一些结果推广到非奇异马尔可夫转移函数的情形。他还研究过概率母函数的有关问题。后来他开始对以无穷维随机微分方程来处理具有无穷自由

度的动力系统发生兴趣，他先研究了基本事实并验证了特殊例子，后来，他习惯于以无穷维的观点来观察即使是有限维的事实，这一习惯引导他将上述问题作为一个在游弋空间中取值的泊松点过程。

伊藤清的成就使人们充分地了解马尔可夫样本道路的无穷小展开。这可以看做是随机领域里的牛顿定律，它提供了起制约作用的偏微分方程和内在的概率机制之间的一个直接转换。它的主要组成部分是布朗运动的函数的微积分。由此而产生的理论是现代概率论(不论是纯粹的还是应用的)基石。伊藤清的工作使人们对工程的计划、控制和最优化，以及其他一些基本上是随机而非确定系统的有关问题和现象，有了深刻的理解。

伊藤清创立随机积分颇负盛名。随机积分是对某些随机过程类适当定义的各种积分的总称，它们在随机过程与随机微分方程的研究和应用中各有其重要作用。以伊藤清的姓氏命名的伊藤积分是对布朗运动定义的一种随机积分。布朗运动的样本函数虽然连续，但几乎所有的样本函数非有界变差，甚至处处不可微，因而无法按样本函数来定义通常的黎曼-斯蒂尔切斯积分或勒贝格—斯蒂尔切斯积分。一般来说，黎曼-斯蒂尔切斯积分定义中的达布和不会以概率 1 收敛到一定的极限，但在适当的条件下，达布和的均方极限存在。伊藤清正是利用这一性质定义了对布朗运动的随机积分。而伊藤清积分最重要的性质是如下所示的著名伊藤公式

$$F(W(b))=F(W(a))+\int_a^b F'(W(t))dW(t)+\frac{1}{2}\int_a^b F''(W(t))d(t),$$

其中 F 是二次连续可微实函数，$W(t)(t\geq 0)$ 是布朗运动。

这个公式及其各种推广在理论上和应用上都有重要的作用。例如，可以用来证明关于布朗运动的鞅刻画的莱维定理。

在概率论与随机过程这个领域中，有不少以伊藤清的姓氏命名的方程、公式、积分、过程等。

伊藤清不但研究成果卓著，而且还培养和造就了整整一代日本的概率论专家。

伊藤清说："科学的目的是从已知推断未知。如果从已得到的资料能作出唯一正确的推断，则可建立确定性模式，分析学为此提供了手段。当现象极其复杂，不可能作唯一推断时，只好从已知来求未知的平均，为此应建立随机性模式，随机分析学为此提供数学手段"。他指出："数学取得了显著的进步，数学各分支相互联系越来越密切，作为有机整体的数学正在形成。此外，与数学有关的其他学科也用了很多高深的数学理论，对作为科学基础的数学的期待是很高的"。

伊藤清的主要专著有：《随机过程论》(1942 年)、《概率论基础》(1944 年)、《论随机微分方程》(1953 年)、《随机过程》(1957 年)、《概率论》(1952 年)等。其中，《概率论》和《随机过程》已译成中文，分别由科学出版社、上海科学技术出版社于 1963 年、1961 年出版。

世界著名的斯普林格出版社 1987 年出版了伊藤清的选集，这部选集差不多是伊藤清科学论文的全集。它反映了伊藤清所作的贡献，主要涉及他所创立的随机微分理论的基础问题，其余论文则讨论扩散理论、布朗运动、回归理论和随机微分方程。该选集所有的论文都反映出伊藤清对概率学科的极为深刻的探讨，并将读者引进了一个重要而又非常活跃的现代数学领域。该选集还有编者所写对伊藤清工作的评论及伊藤清本人评论其研究工作发展的前言。

伊藤清 1981 年以日本数学会理事长的身份曾来我国进行学术访问，并作了学术演讲。

概率论的发展史说明了理论与实际之间的密切关系。许多研究方向的提出，归根到底是有其实际背景的。反过来，当这些方向被深入研究后，又可指导实践，进一步扩大和深化应用范围。概率论作为数理统计学的理论基础是尽人皆知的。下面简略介绍一下概率论本身在各方面的应用情况。

在物理学方面，高能电子或核子穿过吸收体时，产生级联（或倍增）现象，在研究电了—光子级联过程的起伏问题时，要用到随机过程，常以泊松过程、弗瑞过程或波伊亚过程作为实际级联的近似，有时还要用到更新过程（见点过程）的概念。当核子穿到吸收体的某一深度时，则可用扩散方程来计算核子的概率分布。物理学中的放射性衰变，粒子计数器，原子核照相乳胶中的径迹理论和原子核反应堆中的问题等的研究，都要用到泊松过程和更新理论。湍流理论以及天文学中的星云密度起伏、辐射传递等研究要用到随机场的理论。探讨太阳黑子的规律及其预测时，时间序列方法非常有用。

化学反应动力学中，研究化学反应的时变率及影响这些时变率的因素问题，自动催化反应，单分子反应，双分子反应及一些连锁反应的动力学模型等，都要以生灭过程（见马尔可夫过程）来描述。

随机过程理论所提供的方法对于生物数学具有很大的重要性，许多研究工作者以此来构造生物现象的模型。研究群体的增长问题时，提出了生灭型随机模型，两性增长模型，群体间竞争与生尅模型，群体迁移模型，增长过程的扩散模型等。有些生物现象还可以利用时间序列模型来进行预报。传染病流行问题要用到具有有限个状态的多变量非线性生灭过程。在遗传问题中，着重研究群体经过多少代遗传后，进入某一固定类和首次进入此固定类的时间，以及最大基因频率的分布等。

许多服务系统，如电话通信，船舶装卸，机器损修，病人候诊，红绿灯交换，存货控制，水库调度，购货排队等，都可用一类概率模型来描述。这类概率模型涉及的过程叫排队过程，它是点过程的特例。排队过程一般不是马尔可夫型的。当把顾客到达和服务所需时间的统计规律研究清楚后，就可以合理安排服务点。

在通信、雷达探测、地震探测等领域中，都有传递信号与接收信号的问题。传递信号时会受到噪声的干扰，为了准确地传递和接收信号，就要把干扰的性质分析清楚，然后采取办法消除干扰。这是信息论的主要目的。噪声本身是随机的，所以概率论是信息论研究中必不可少的工具。信息论中的滤波问题就是研究在接收信号时如何最大限度地消除噪声的干扰，而编码问题则是研究采取什么样的手段发射信号，能最大限度地抵抗干扰。在空间科学和工业生产的自动化技术中需要用到信息论和控制理论，而研究带随机干扰的控制问题，也要用到概率论方法。

概率论进入其他科学领域的趋势还在不断发展。值得指出的是，在纯数学领域内用概率论方法研究数论问题已经有很好的结果。在社会科学领域，特别是经济学中研究最优决策和经济的稳定增长等问题，也大量采用概率论方法。正如拉普拉斯所说：“生活中最重要的问题，其中绝大多数在实质上只是概率的问题”。

第 4 章　随机变量的分布产生的背景

在自然界和现实生活中，一些事物都是相互联系和不断发展的。在它们彼此间的联系和发展中，根据它们是否有必然的因果联系，可以分成截然不同的两大类：一类是确定性的现象。这类现象是在一定条件下，必定会导致某种确定的结果。另一类是不确定性的现象。这类现象是在一定条件下，它的结果是不确定的。正因为这样，我们在这一类现象中，就无法用必然性的因果关系，对个别现象的结果事先做出确定的答案。事物间的这种关系是属于偶然性的，这种现象叫做偶然现象，或者叫做随机现象。随机现象从表面上看，似乎是杂乱无章的、没有什么规律的现象。但实践证明，如果同类的随机现象大量重复出现，它的总体就呈现出一定的规律性。大量同类随机现象所呈现的这种规律性，随着我们观察的次数的增多而愈加明显。比如掷硬币，每一次投掷很难判断是那一面朝上，但是如果多次重复的掷这枚硬币，就会越来越清楚地发现它们朝上的次数大体相同。我们把这种由大量同类随机现象所呈现出来的集体规律性，叫做统计规律性。随机现象统计规律的研究，可以转化为实变量的实值函数的研究。通过引进随机变量及其分布函数的概念，实现这一转化。一般的统计类的书籍只是简要介绍各类分布的定义、性质和应用。而对于其产生的背景，即它是怎么来的，为什么要研究这类分布？没有一个详细的讲解。针对这种情形在查阅了大量中、外文资料后，对各类分布的产生背景进行了整理汇编，并用统一的符号进行叙述，介绍各类分布产生的背景。

4.1 离散型随机变量的分布产生的背景

伯努利试验及其分布只有两种可能结果的随机试验，称为 Bernoulli 试验。这是一类特别常见的试验，例如射击的中与不中，抽取产品的合格与不合格等。都是两种结果，若将一个随机试验独立重复（试验的条件不变）地做 n 次，则简称为 n 重随机试验，这样可定义 n 重伯努利试验。

1.伯努利分布

在一个伯努利实验中，若令 $X=\begin{cases}0, & \text{试验失败}\\ 1, & \text{试验成功}\end{cases}$，称此随机变量 X 为伯努利计数变量。

若记 $p=P(X=1)>0$，$q=1-p>0$，则其分布列为

$$\begin{pmatrix}0 & 1\\ q & p\end{pmatrix},$$

称这种概型为伯努利概型，此分布 $\{p,q\}$ 为伯努利分布或 0-1 分布。

2. n 重伯努利试验及其产生的分布——二项分布

我们现在用事件独立性来研究，如果我们一次抛掷 n 枚相同的硬币，要求“恰好出现 k 个正面”这一事件的概率 $P_n(k)$，这样一个“一次抛掷 n 枚相同硬币”的随机试验，可

以用另一种等价的方式来进行，每次抛掷一枚硬币，共抛掷n次，容易理解，这n次抛掷的结果是相互独立的，因而如果把相同条件下抛掷一枚硬币看作是一次试验，就意味着这n次试验是相互独立的，即：试验的结果是相互独立的，一般的说，试验E只有两个可能的结果

$$A\text{及}\ \overline{A},$$

并且

$$P(A)=p,P(\overline{A})=1-p=q,$$

其中$0<p<1$，把E独立地重复n次的试验构成n重伯努利试验，并记作E^n。

上述“一次抛掷n枚相同硬币”的试验就可以看作是一个n重伯努利试验。一个伯努利试验的结果可以记作

$$W=(W_1,W_2,\cdots,W_n),$$

其中的$W_i\ (1\le i\le n)$或者为A或者为$\overline{A}$，因而这样的W共有2^n个，它们的全体就是这伯努利试验的样本空间Ω。

对于$W=(W_1,W_2,\cdots,W_n)\in\Omega$，如果$W_i\ (1\le i\le n)$中有$k$个为$A$，则必有$n-k$个为$\overline{A}$，于是由独立性，即得

$$P(W)=p^k q^{n-k}$$

记$B_k=\{\ n$重伯努利试验中事件A出现k次$\}$，则由概率的有限可加性即得：

$$P(B_k)=\sum p(W),$$

对于$W\in B_k$，已知

$$P(W)=p^k q^{n-k},$$

而B_k中这样的W共有C_n^k 个，所以有

$$P(B_k)=C_n^k p^k q^{n-k}(0\le k\le n)$$

令“ξ”表示n重伯努利试验中事件A出现的次数，则

$$P_k=P(\xi)=P(B_k)=C_n^k p^k q^{n-k}(0\le k\le n)$$

注意到$P_k=C_n^k p^k q^{n-k}(0\le k\le n)$恰好是二项式$(p+q)^n$的展开式中的第$k+1$项，由此人们给分布列$P_k=C_n^k p^k q^{n-k}(0\le k\le n)$起了一个名字，称它为二项分布，记为

$$b(k;\ n,\ p)=C_n^k p^k q^{n-k}(0\le k\le n)$$

一个随机变量的分布列如果是二项分布，也称该随机变量服从二项分布。

3.几何分布（可列重伯努利试验产生的分布）

服从二项分布的随机变量只取有限个值，而服从几何分布的随机变量取可列个值。在一个伯努利试验中，每次试验成功的概率为p，失败的概率为

$$q=1-p(0<p<1)$$

设试验进行到第ξ次才出现成功，则

$$P(\xi=k)=pq^{n-k},k=1,2,\cdots$$

容易看到$pq^{n-k}\ (k=1,2,\cdots)$是几何级数$\sum_{k=1}^{+\infty}pq^{k-1}$的一般项，于是人们称它为几何分布，

并记

$$Ge(k,p)=pq^{k-1}$$

几何分布具有无记忆性，在已经试验n次尚未成功的概率，与重新开始试k次未成功的概率相等，而与n无关。

即当$X\sim Ge(k,p)$时，$P(\xi>n+k>n)=P(\xi>k)$。反之，有无记忆性的离散型分布，必为几何分布。

几何分布是一种“等待分布”我国古代寓言“守株待兔”里盼望兔子撞树的等待时间，一些电子元件的寿命一个自动化控制系统第一次出现某个控制命令的等待时间，网络上某个网站第一次发出某类信息的时间等，按离散的时间计数可以认为是几何分布。

4.负二项分布

伯努里试验无限地进行下去，但要试验成功$r\,(r\geq 1)$次才停止试验，仍用ξ表示，必须进行试验的次数，则$\xi=k$时，第k次一定是成功，而前$k-1$次成功$r-1$次，从而至第k次累计成功r次，故

$$p_k\,def\,P(\xi+k)=pC_{k-1}^{r-1}p^{r-1}q^{k-1-(r-1)}=C_{k-1}^{r-1}p^rq^{k-r}\,(k=r,r+1,\cdots),$$

有如上形式的分布$\{P_k\}$，称为参数为p和r的负二项分布，也称为 Pascal 分布，相应的随机变量X服从负二项分布，记$X\sim F(r,p)$。

由定义知，负二项分布也是一个离散型的等待分布，它是第r个质点（故障、命令以及信息等）到达的离散型时刻，且显然$F(1,p)=Ge(k,p)$。

5.泊松（Poisson）分布

历史上，泊松分布是作为二项分布的近似，由法国数学家泊松于 1837 年引入，近数十年来，泊松分布日益显示其重要性，成了概率论中最重要的几个分布之一，观察某公共汽车站在单位时间里来站乘车的乘客数，宇宙中单位体积内星球的个数、耕地上单位面积内杂草的数目等，如果相应的变量用ξ表示，那么实践表明，ξ的统计规律近似地为

$$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots,$$

其中$\lambda>0$是常数，这个分布称作是参数为λ的泊松分布，记作：$P(k,\lambda)$。

下面我们介绍一下泊松流。源源不断到来的质点，称作流，这里质点可以是空间粒子、商场顾客、路口的车辆，从而形成粒子流、顾客流、车流等，一项项需求、一条条信息或指令也构成流。

定义：设某个流在$(0,t]$时段内来到的质点数为$\xi_{(0,t]}$，并记$P_k(t)=P(\xi_{(0,t]}=k)$，这里$k\in Z^+def\{0,1,2,\cdots\}$，$t\geq 0$，称这个流为泊松流。

如其满足下面四个条件：

(1)增量独立性，设Δ_1和Δ_2是两个不相交的时间段，则在这两个时段内所增加的质点事件是独立的，即两事件$(\xi_{\Delta_i}=k_i)$，$i=1,2$独立，这里k_1和k_2是两个任意的非负整数。

(2)增量平稳性，$P(\xi_{(s,s+t]}=k)=P(\xi_{(0,t]}=k)=P_k(t)$，$k\in Z^+,0\leq s$。

(3)有限性，在任意有限长的时段内只来有限多个质点，即

$$P(\xi_{(0,t]}=\infty)=0,\forall t\geq 0,$$

或

$$\sum_{k=0}^{\infty} P_k(t)=P(\xi_{(0,t]}<\infty)=1$$，

且设 $P_0(t)$ 不恒为 1。

(4)普通性，$\sum_{k=2}^{\infty} P_k(t)=P(\xi_{(0,t]}\geq 2)=o(t)$，即认为在任一瞬间来两个以上质点是可以忽略不计的。增量独立性中 k_1 和 k_2 可以不同，而增量平稳性要求是同一个 k，平稳性说明在长为 t 的时段内来质点数的概率规律。

从定义上看，泊松流的条件好像很苛刻，其实泊松流的应用是很广泛的，生产中和社会生活中很多流都可用它来近似，如果网站在 $(0,t]$ 时段内收到的点击数，电话交换台收到的呼叫数，交通枢纽的客流和车流，自动化到的寻呼流（次数），网络通信中收到某类信号的次数等，常常都可用泊松流刻划。为了说明泊松分布产生的背景，先介绍两个定理。

定理 1：设 $X_n \sim B(n,P_n)$，即对固定的几次试验中，每次试验成功概率是 P_n，又设存在极限 $\lim_{n\to\infty} nP_n=\lambda>0$，则对任意非负整数 k，有

$$P(X_n=k)=C_n^k P_n^k(1-P_n)^{n-k}\xrightarrow{n\to+\infty}\frac{\lambda^k}{k!}e^{-\lambda}$$，

即泊松逼近定理。

定理 2（泊松定理）：设 $\xi_{(0,t]}$，$t\geq 0$ 是泊松流，则存在某正数 λ，使

$$P_k(t)=P(\xi_{(0,t]}=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}\ (k=0,1,2,\cdots)$$

由泊松逼近定理知泊松分布可以看作是二项分布的近似，现在由定理 2 知 $\xi_{(0,t]}\sim P(\lambda,t)$，即：泊松流有参数为 λt 的泊松分布，在上式中取 $t=1$（单位时间），即得到参数为 λ 的泊松分布，取 $t=2$，则得到参数为 2λ 的泊松分布，因此泊松流是产生泊松分布的直接且最重要的背景。

6.超几何分布

假设一个盒子里含 b 个蓝色的弹子和 r 个红色的弹子，作一个实验，包含几次试验，每次随机地选取一个弹子，观察它的颜色，然后把该弹子放回盒里，这类实验常归属于有放回抽样，在这样的情形里，若 ξ 是一个随机变量，它表示在几次试验中选中蓝色的弹子的数目，则利用二项分布式，我们知道成功数 k 的确切概率是

$$P(\xi=k)=C_n^k\frac{b^k r^{n-k}}{(b+r)^n}\ (\ k=0,1,2,\cdots\ ),$$

此时 $p=\dfrac{b}{b+r}$，$q=1-p=\dfrac{r}{b+r}$。

若为了无放回抽样而修改以上的叙述，即选取弹子后不放回盒里，则

$$P(\xi=k)=C_n^k\frac{b^k r^{n-k}}{(b+r)^n},$$

这是超几何分布。

4.2 连续型随机变量的分布产生的背景

1.均匀分布 $U[a,b]$ 及其产生的背景

在估计计算及测量引起的随机误差问题中，有一类误差具有“均匀性”，相应的分布叫做均匀分布。

其定义如下：称连续型随机变量 X 服从区间 $[a,b]$ 上的均匀分布，并记为 $X\sim U[a,b]$，如其概率密度函数有如下形式

$$f_X(x)=\begin{cases}\dfrac{1}{b-a},x\in[a,b]\\ 0,x\notin[a,b]\end{cases},$$

则它的分布函数为

$$F_X(x)=\begin{cases}0,x<a\\ \dfrac{x-a}{b-a},a\le x<b\\ 1,x\ge b\end{cases}$$

仿上可定义在开区间及半开半闭区间上的均匀分布，从概率规律看来，$U[a,b]$ 与 $U(a,b)$ 及 $U[a,b)$ 没有实质性差别。

回顾概率密度函数的概率意义，$f(x)dx$ 可视为 X 在 x 点附近（微分邻域）的概率，故 $[a,b]$ 上的均匀分布表明 X 在 $[a,b]$ 上每一点附近都是等概率的，如将此区间四等分，那么 x 落在每一个小区间的概率都是 1/4，变成古典概型问题，这两类概型的共同点是均匀性，区别是概率空间的无限（且不可列）和有限，如将前述的每一个小区间再四等分，则 x 落在每一个细分了的小区间的概率都是 1/16，如此继续，可以得到一列古典概型，反之，均匀分布概型可以看成古典概型的一个从有限到无限的极端情形，这也可以看成是均匀分布的一个产生背景。

产生均匀分布的另一个背景是舍入误差问题，在计算机上计算，对实数要作近似处理，所谓单精度实数和双精度实数，就是这类处理的规则和结果，设将实数小数点后第 6 位四舍五入，则舍入误差 X 在 $(-0.5\times10^{-5},0.5\times10^{-5})$ 内任意一点的（微分）领域应为等可能的，因此可以为 X 服从此区间上的均匀分布。

2.指数分布与 Γ 分布的定义与产生的背景

设 $\xi_{(0,t]}$，$t\ge0$ 是泊松流，强度为 λ，则由泊松流的泊松定理，有

$$P_k(t)=P(\xi_{(0,t]}=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad k=0,1,2,\cdots$$

令 η 为泊松流中第一个质点到达的时刻，由于

$$(\eta>t)=(\xi_{(0,t]}=0),\forall t>0,$$

故由泊松定理，

$$P(\eta>t)=e^{-\lambda t},$$

从而

$$F_\eta(t)=1-e^{-\lambda t},$$

又当$t \le 0$时，显然有

$$F_\eta(t)=0$$

求导得到泊松流中第一个质点到达时刻η的概率密度函数，

$$f_\eta(t)=\begin{cases}\lambda e^{-\lambda t}, t>0\\ 0, t\le 0\end{cases}, \quad (4\text{–}1)$$

这里的参数正是泊松流中的强度λ.

若令η_r为泊松流中第r个质点到达的时刻，由于

$$(\eta_r>t)=(\xi_{(0,t]}<r), \forall t>0,$$

根据泊松定理，有

$$P(\eta_r>t)=\sum_{k=0}^{r-1}\frac{(\lambda t)^k}{k!}e^{-\lambda t}$$

因此

$$f_{\eta_r}(t)=\frac{d}{dt}F_{\eta_r}(t)=\frac{d}{dt}[1-P(\eta_r>t)]=-\sum_{k=1}^{r-1}\frac{\lambda(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}+\sum_{k=0}^{r-1}\frac{(\lambda t)^k}{k!}\lambda e^{-\lambda t}$$

$$=\frac{\lambda(\lambda t)^{r-1}}{(k-1)!}e^{-\lambda t}$$

当$t \le 0$时，显然有

$$f_{\eta_r}(t)=0,$$

故

$$f_{\eta_r}(t)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}, t>0\\ 0, t\le 0\end{cases}, \quad (4\text{–}2)$$

其中Γ函数$\Gamma(r)=(r-1)!$。

定义：设连续型随机变量X的分布密度$f(x)$有形式 4–1，其中$\lambda>0$为常数，则称X服从参数为λ的指数分布，记作$X\sim E(\lambda)$。

如其概率密度函数有形式 4–2，则称X服从参数为r和λ的Γ分布，记作$X\sim\Gamma(r,\lambda)$，其中r及λ分别叫做形状参数及尺度参数。

在可靠性理论的一些产品设备和系统，例如电子元件芯片，出现第 1 个故障的时刻，可以认为是指数分布的而第r个故障出现的时刻服从Γ分布。指数分布由其产生的背景可知，也是一种等待分布。连续时间的“守株待兔”的例子，待兔的时间便是服从指数分布的随机变量，对应的离散型的等待分布是几何分布；而分布$\Gamma(r,\lambda)$对应的离散型的分布是负二项分布$F(r,p)$。

3.贝塔分布

概率论中有一种称为贝塔（β，beta）分布的概率密度分布函数。它的数学形式是

$$f(x)=\frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}(0<x<1,p>0,q>0),$$

其中变量x仅能出现于 0 到 1 之间，p,q是两个大于 0 的参数。$B(p,q)$的含义是

$$B(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}dx,$$

它与Γ函数，有如下关系

$$B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},$$

而我们介绍过的阶乘符号!与Γ的关系是

$$n!=\Gamma(n+1),$$

所以贝塔分布也可以写为

$$f(x)=\frac{(m+n+1)!}{m!n!}x^{m}(1-x)^{n}$$

现在考虑从最复杂原理加适当的约束条件推求这个概率密度分布函数的问 题。根据过去的经验，容易看出它可能是下面两个约束条件与最复杂原理的应用结果。

变量x的对数的平均值为固定值（等价于几何平均值为常数）

$$u=\int_0^1(\ln x)f(x)dx,$$

（$1-x$）的对数的平均值也是固定之值

$$v=\int_0^1[\ln(1-x)]f(x)dx,$$

作为概率密度，当然还有

$$\int_0^1 f(x)dx=1,$$

根据上面的三个约束公式和最复杂原理，利用拉哥朗日方法，构造的F函数是

$$F=\int_0^1-f\ln f dx-C_1(\int_0^1 f dx-1)+C_2(\int_0^1\ln x f dx-u)+C_3(\int_0^1\ln(1-x)f dx-v),$$

求F对未知的概率密度f的偏微商，并且令它等于0（利用了最复杂原理），我们得到

$$f(x)=e^{C_1-1}x^{C_2}(1-x)^{C_3},$$

利用分布函数的积分应当等于1的约束和积分知识我们得到

$$e^{C_1-1}=\frac{1}{B(C_2+1,C_3+1)},$$

所以分布函数可以写为

$$f(x)=\frac{1}{B(C_2+1,C_3+1)}x^{C_2}(1-x)^{C_3},$$

显然，这个公式的外形已经与贝塔分布一致了。余下的问题是利用关于u，v的约束公式可以求出C_2，C_3。使这个公式通过u，v来表示。由于u，v与C_2，C_3的关系比较复杂，我们没有得到具体的关系式。但是概率密度分布函数的形状与概率论中的贝塔分布一致就已经达到了我们的目的：介于0~1之间的变量的两种几何平均值固定和最复杂原理相结合可能是一些贝塔分布形成的原因。

4.Gamma 分布

连续型的随机变量x（或者说一个广义集合的标志变量）如果它的概率密度分布函数

$f(x)$ 符合

$$f(x)=\frac{\beta^n}{(n-1)!}x^{n-1}e^{-\beta x},x>0 \tag{4-3}$$

这个概率密度函数称为伽码（Γ.Gamma）分布。它也是著名的皮尔逊概率分布函数簇中的重要一员，称为皮尔逊Ⅲ型分布。它的曲线有一个峰，但左右不对称。在自然界中服从这种分布的现象不少。

公式中有两个参数 n，β。由于这种分布对自变量要求有一个大于等于零下限，拟合资料时又比正态分布的弹性大，在我国的水文界广泛用皮尔逊Ⅲ型分布来模拟水文数据系列。中国新规范规定："频率曲线型一般应采用 P-Ⅲ型分布，经分析论证后可采用其他线形。我们现在利用最复杂原理寻找形成这种分布的物 理原因。如果分析这个公式的外型，不难发现它既具有负指数分布中存在的指数部分，也存在幂分布公式中的幂函数特点。我们记得指数分布对应着标志变量的代数平均值不变的约束，而幂函数对应着变量的几何平均值不变的约束。于是容易猜想到 Gamma 分布的约束条件就是变量的代数平均值和几何平均值都是固定值。确实，一个必然大于零的随机变量（如河水的流量）其代数平均值和几何平均值分别为固定值（不同的），并且它出现什么值的不确定性（结局的复杂性）最大，不难利用与前面类似的方法推导出它的概率分布函数必然是 Gamma 分布。

用 $f(x)$ 表示随机变量 x 的概率密度分布函数，有

$$\int_0^{+\infty} f(x)dx=1, \tag{4-4}$$

以 u 代表变量的代数平均值，既有

$$u=\int_0^{+\infty} xf(x)dx \tag{4-5}$$

根据前面对几何平均值的讨论，我们也可以把几何平均值不变写为变量的对数的代数平均值不变，这个约束可以写为 v 不变，这里

$$v=\int_0^{+\infty}(\ln x)f(x)dx, \tag{4-6}$$

而变量的信息熵（对应于复杂程度）为

$$H=-\int_0^{+\infty} f(x)\ln f(x)dx,$$

在式 4-4、4-5、4-6 的约束下，让熵最大反求分布函数时，用拉格朗日方法构造的函数 F 为

$$F=-\int_0^{+\infty} f\ln fdx+C_1(\int_0^{+\infty} fdx-1)+C_2(\int_0^{+\infty} xfdx-u)+C_3(\int_0^{+\infty}\ln xfdx-v),$$

这里的 C_1,C_2,C_3 是待定的常数。熵 H 最大（最复杂原理），也就是 F 最大，将上式对 F 求偏微商，F 最大就是它的偏微商为 0，于是得到

$$f(x)=e^{C_1-1+C_2x+C_3\ln x},$$

经过整理，可以就得到

$$f(x)=e^{C_1-1}x^{C_3}e^{C_2x},$$

这个结果就是求得的分布函数。由于各个 C_i 都是一些常数，所以这个公式说 明分布函数

是幂函数与指数函数的乘积。它的外型与本节最初给出的 Gamma 分布公式已经相同了。

在一个广义集合（客观事物、系统、抽样实验）中如果变量（标志值）的代数平均值和几何平均值是不变的，而其复杂程度（熵）最大，那么各个个体为各种标志值（变量的各种取值）的概率（占的百分比）必然是 Gamma 分布（皮尔逊Ⅲ型分布）。

这样我们就利用最复杂原理（最大熵原理）说明了 Gamma 分布的物理成因。

5.负指数分布

现在我们具体地把约束条件限定为下面的两个

$$\int_0^{+\infty} f(x)dx=1 ,\qquad (4\text{—}7)$$

$$\int_0^{+\infty} xf(x)dx=a ,\qquad (4\text{—}8)$$

与此约束条件对应的函数显然是

$$u_1=1,$$

$$u_2=x,$$

我们把这两个具体函数代入

$$f(x)=e^{-1+\sum C_i u_i(x)},$$

（满足最复杂原理和i个约束关系的广义集合的分布函数）中，也就得到了具体满足这两个约束并且符合最复杂原理的广义集合的分布函数，即有

$$f(x)=e^{-1+C_1x+C_2x},$$

目前上式中的两个常数C_1，C_2的值还不知道。但是利用式 4-7、4-8 这两个关系容易得到

$$f(x)=\frac{1}{a}e^{-\frac{x}{a}} \qquad (4\text{–}9)$$

于是得到结论：广义集合的分布函数（相对密度形式的）如果应当满足式 4-7 和 4-8 的关系，并且其复杂程度最大，那么其分布函数（相对密度）就必然是公式 4-9，即它是一个负指数函数。

6.威伯（Weibull）分布

威布尔分布是 Weibull 在拟合合金钢的强度中提出来的，因而得名。在可靠性问题中是一个最重要的分布。

以连续变量x为标志的广义集合，（即随机变量为x），如果已经知道变量的n次方的代数平均值为固定值u，就可以写为

$$u=\int_0^{+\infty} x^n f(x)dx$$

我们还知道变量的对数的平均值为固定值（等价于几何平均值为常数），就有

$$v=\int_0^{+\infty}(\ln x)f(x)dx ,$$

上面约束条件公式中的$f(x)$仍然是概率密度分布函数，而且从第二个约束知 道变量必须是大于零的变量（只能取正值，注意积分下限）。

对于概率密度分布函数，当然有

$$\int_0^{+\infty} f(x)dx=1$$

如果该广义集合具有随机性，它应当遵守最复杂原理，根据上面的三个约束公式，利用拉格朗日方法求分布函数时，构造的 F 函数是

$$F=-\int_0^{+\infty} f\ln f dx+C_1\left(\int_0^{+\infty} f dx-1\right)+C_2\left(\int_0^{+\infty}\ln x f dx-v\right)+C_3\left(\int_0^{+\infty}x^n f dx-u\right)$$

求 F 对未知的概率密度 f 的偏微商，并且令它等于 0（利用了最复杂原理）， 得到

$$\ln f(x)=(C_1-1)+C_2\ln x+C_3x^n,$$

整理得

$$f(x)=Cx^{C_2}e^{C_3x^n},$$

在概率论中我们已经知道威伯分布（Weibull）就是具有下面形状的概率分布函数

$$f(x)=\frac{n}{u}x^{n-1}e^{-\frac{x^n}{u}} \tag{4—10}$$

我们看到用给定约束和最复杂原理推导的分布函数与概率论中的威伯分布具有相同的函数形状。这也就说明威伯分布的形成的物理原因有可能（不见得全部）就是客观事物的随机性（伴随着最复杂原理有效）和这里给出的约束条件。公式 4-10 中的参数 u，也就是变量的 n 次方的平均值。在实际观测中，对于必然是正值的变量 x，它一般具有一个大于零的最小值 a，那么威伯分布就写为

$$f(x)=\frac{n}{u}(x-a)^{n-1}e^{-\frac{(x-a)^n}{u}}\ (0<a\le x),$$

从外型上看，这个分布函数有三个参数，其中 a 描述了变量的最小正值（下限）， n 描述了多少次的方，u 是变量 x 减 a 的 n 次方的平均值。

以上是我们从最复杂原理角度认识威伯分布的思路。

7.瑞利（Rayleigh）分布

当威伯分布中的参数 $n=1$ 时，就得到了负指数分布。所以负指数分布也是威伯分布的一个特例。在威伯分布中的参数 $n=2$ 时，又得到了所谓瑞利分布

$$f(x)=\frac{2(x-a)}{u}e^{-\frac{(x-a)^2}{u}},$$

所以瑞利分布也是威伯分布的一个特例。

8.对数正态分布

所谓对数正态分布是指变量 x 的概率密度的分布函数 $f(x)$ 满足下面的关系

$$f(x)=\frac{1}{\sqrt{2\pi}cx}e^{-\frac{(\ln x-a)^2}{2\sigma^2}}\quad (x>0)$$

这个公式与正态分布公式类似。但是系数部分的分母上多了自变量 x，另外， 在指数部分的 x 处变成了该变量的对数值了。这里的常数 a，σ 的含义有变化。a 是变量 x 的对数的平均值，σ 是变量 x 的对数的标准差。由于对数仅能用于正实数，所以变量 x 必须是大于零的数。$f(x)$ 就是变量 x 出现于 $x-0.5$ 到 $x+0.5$ 范围的概率密度。现在用最大熵

原理配合约束条件推导这个公式。

约束条件为

$$\int_0^{+\infty} f(x)dx=1, x>0$$

，

一个新的约束是变量的对数的平均值 a 是固定值

$$a=\int_0^{+\infty}(\ln x)f(x)dx, x>0$$

，

另外一个类似正态分布的约束条件就是变量的对数的标准差为固定值，即有

$$\sigma^2=\int_0^{+\infty}(\ln x-a)^2 f(x)dx, x>0$$

，

根据拉格朗日方法，定义一个新函数 F

$$F=-\int_0^{+\infty} f\ln f dx+C_1\left(\int_0^{+\infty} f dx-1\right)+C_2\left(\int_0^{+\infty}\ln x f dx-a\right)+C_3\left(\int_0^{+\infty}(\ln x-a)^2 f dx-\sigma^2\right)$$

，

上面公式中的 f 就是 $f(x)$。

对最复杂原理的利用体现在上式对 f 的偏微商应当等于 0，于是得到

$$f(x)=e^{C_1-1}x^{C_2}e^{C_3(\ln x-a)^2}$$

，

利用这些约束关系可以解出三个常数 C_1，C_2，C_3，经过整理就得到对数正态分布公式。

上面的推证是在变量必须大于 0 的必然要求下进行的。如果变量不仅是大于 0，而且是大于某个正数 b，上面各个公式中的积分下限就改为 b。这实际是把坐标向右移动了 b。

与此对应的概率密度公式是

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma(x-b)}e-\frac{(\ln(x-b)-a)^2}{2\sigma^2};$$

这样我们就利用最复杂原理配合这些新的约束条件得到了概率论中所谓的对数正态分布函数。

4.3 高斯与误差正态分布

最小二乘法是提供“观测组合”的主要工具之一，它依据对某事件的大量观测而获得“最佳”结果或“最可能”表现形式。如已知两变量为线性关系 $y=ax+b$，对其进行 n（$n>2$）次观测而获得 n 对数据。若将这 n 对数据代入方程求解 a,b 之值则无确定解。最小二乘法提供了一个求解方法，其基本思想就是寻找“最接近”这 n 个观测点的直线。最小二乘法不仅是 19 世纪最重要的统计方法，而且还可以称为数理统计学之灵魂。相关回归分析、方差分析和线性模型理论等数理统计学的几大分支都以最小二乘法为理论基础。作为其进一步发展或纠正其不足而采取的对策，不少近现代的数理统计学分支也是在最小二乘法基础上衍生出来的。正如美国统计学家斯蒂格勒(S.M.Stigler)所说，“最小二乘法之于数理统计学犹如微积分之于数学”。

最小二乘法创立的历史过程充满着丰富的科学思想，这些对今日的数学创造仍有着重要的启示意义。本文旨在全面认识最小二乘法的历史系统发育过程以及创立者的思路。

1.先驱者的相关研究

天文学和测地学的发展促进了数理统计学及其他相关科学的发展。丹麦统计史家哈尔德曾指出天文学在数理统计学发展中所起的作用。“天文学自古代至 18 世纪是应用数学中最发达的领域。观测和数学天文学给出了建立数学模型及数据拟合的最初例子，在此种意义下，天文学家就是最初的数理统计学家。天文学的问题逐渐引导到算术平均，以及参数模型中的种种估计方法，以最小二乘法为顶峰”。这也说明了最小二乘法的显著地位。

有关统计计算思想记载的著作要首推天文学家罗杰柯茨的遗作，即 1715 年其所发论文中所蕴含的统计方法，亦即对各种观测值赋予加权后求其加权平均。尽管当时得到认可，然而事实证明如此计算的结果不太精确。

1749 年，欧拉(L.Euler，1707~1783)在研究木星和土星之间相互吸引力作用对各自轨道影响时，最后得到一个含 8 个未知量 75 个方程的线性方程组。欧拉的求解方法繁杂而奇特，只能看作是一次尝试。

1750 年，天文学家梅耶(T.Meiyer，1723~1762)通过对月球表面上某定点的观测，得到一含 3 个未知数 27 个方程的线性方程组。以其中一个方程系数为准，按各方程中此系数的大小分组，较大的 9 个、较小的 9 个和剩下的 9 个分别组成一组。每组内的 9 个方程相加，得到一个方程。由得到的 3 个方程而求解 3 个未知数。梅耶认为，如此所得解之误差比任意选 3 个方程而求解之误差要小得多，仅为其 3/27=1/9(实际上应为 3/9 = 1/3)。由此他得出解类似方程组的一套系统方法，并曾一度相当流行。直到 1760 年，罗杰博斯科维奇(1711~1787)在研究地球真实形状的有关问题时才指出其不足。他认为梅耶确定方程组解的方法还不够精确，应充分满足实际准则，其中包括把一组观测值代入方程组时所产生误差的绝对值之和极小化准则。1787 年，拉普拉斯在研究天文学时，得到一个含有 4 个未知数 24 个方程的线性方程组。其求解方法与梅耶相似，先把 24 个方程编号，然后得出 4 个方程，以便解出 4 个未知数。这 4 个方程依次为：第一个方程，24 个方程之和；第二个方程，前 12 个方程之和减去后 12 个方程之和；第三个方程，编号为 3，4，10，11，17，18 的方程之和减去编号 1，7，14，20 的方程之和；第四个方程，编号为 2，8，9，15，16，21，22 的方程之和减去编号为 5，6，12，13，19 的方程之和。拉普拉斯并没有给出如此组合的原因，但可以看到如此组合可使同一个方程至少被使用两次，而前 22 个方程被使用三次。这已比前述结果前进了一大步。由此可见，早期的数学家们致力于组合方程而忽视了整体的均衡性。纵观数学史，在每一新的理论创立之前，总是离不开先驱者的努力。这些先驱者的见解往往不完美，甚至含有漏洞和缺陷，但他们的工作对于新思想、新理论的创立是十分必要的。

2.勒让德创立最小二乘法

勒让德(A.M.Legendre，1752~1833)是法国军事学校的教授，曾任多届政府委员，后来成了多科工艺学校的总监，直至 1833 年逝世。他一直保持热情而有规律的数学研究工作，由于解决了许多类型的问题，其名字常存于许多定理之中。数学史家克莱因(M.Kline，1908~1992)认为勒让德之所以名列拉格朗日(J.L.Lagrange，1736~1813)、拉普拉斯、蒙日(G.Monge，1746~1818)之后，是因为其工作不如这三人深刻。尽管勒让德的工作引起许多

重要理论的产生，但这只是在他的研究成果被更强有力的思想接受后才实现的。最小二乘法就是一个典型实例。

最小二乘法最早出现在勒让德 1805 年发表的论著《计算彗星轨道的新方法》附录中。该附录占据了这本 80 页小册子的最后 9 页，在前面关于卫星轨道计算的讨论中没有涉及最小二乘法，可以推测他当时感到这一方法尚不成熟。勒让德在该书 72~75 页描述了最小二乘法的思想、具体做法及其优点。以引进这种方法的理由为开端："所研究的大多数问题都是由观测值来确定其结果，但这几乎总产生形如 $E=a+bx+cy+fz+\cdots$ 方程的方程组，其中 $a,b,c,f,\cdots$ 是已知系数，它们从一个方程到另一个方程是有变动的。$x,y,z,\cdots$ 是未知的，它们必须根据将每个方程 E 化为 0 或很小的量来确定"。用现代术语可描述为，一 n 个未知量 m 个方程的线性方程组($m>n$)，

$$E_j=a_{j0}+a_{j1}x_1+a_{j2}x_2+\cdots+a_{jn}x_n\ (j=1,2,\cdots,m)$$

寻找"最佳"近似解，以使所有 E_j 都变小。勒让德认为："赋予误差的平方和为极小， 则意味着在这些误差间建立了一种均衡性， 它阻止了极端情形所施加的过分影响。这非常好地适用于揭示最接近真实情形的系统状态"。

为了确定误差平方的最小值， 勒让德运用了微积分工具。即为使平方和

$$\sum_{j=1}^{m}E_j^2=E_1^2+E_2^2+\cdots+E_m^2,$$

在 x_i 变动时有最小值，则它对 x_i 的偏导数必为 0。由此得如下线性方程组

$$\sum_{j=1}^{m}a_{ji}a_{j0}+x_1\sum_{j=1}^{m}a_{ji}a_{j1}+\cdots+x_n\sum_{j=1}^{m}a_{ji}a_{jn}=0(i=1,2,\cdots,n),$$

这样，就得到一含有 n 个未知量 n 个方程的线性方程组，用"现成的方法"是可以解出的。

关于最小二乘法的优点，勒让德指出以下几条：

(1)通常的算术平均值是其特例。即 $n=1,a_{j1}=-1$ 时，令 $b_j=a_{j0}$，则误差的平方和为

$$(b_1-x)^2+(b_2-x)^2+\cdots+(b_m-x)^2,$$

对其求关于 X 的偏导数， 则使此和极小的方程是

$$(b_1-x)+(b_2-x)+\cdots+(b_m-x)=0,$$

故解为

$$x=\frac{b_1+b_2+\cdots+b_m}{m},$$

它正是 m 个观测值的算术平均值。

(2)如果观测值全部严格符合某一方程组的要求，则此解必是最小二乘法的解。

(3)如果舍弃或增加观测值，则修改所得方程组即可。

勒让德的成功在于他从一个新的角度来看待这个问题，不像其前辈那样致力于找出几个方程(个数等于未知数的个数)再去求解，而是考虑误差在整体上的平衡。从某种意义讲，最小二乘法是一个处理观测值的纯粹代数方法。要将其应用于统计推断问题就需要考虑观测值的误差，确定误差分布的函数形式。

3.随机误差的早期研究

天文学家伽利略(G.Galileo，1564~1642)可能是第一个提出随机误差概念并对其有所研

究的学者。 他在 1632 年出版的著作《关于两个主要世界系统的对话》中提及这个问题。尽管他用观测误差这个名称，但所描述的性质实则为现在的随机误差分布。伽利略认为：

(1)所有观测值都可能有误差，其源于观测者、仪器工具及观测条件等。

(2)观测误差对称的分布在 0 的两侧，因仪器工具使得观测值比真值大或小的可能性是等同的。

(3)小误差出现的频率大于大误差。

由此可见，伽利略所设想的误差分布函数 $f(x)$ 应满足关于 y 轴对称，且随 $|x|$ 增加而递减等条件。这个定性式讨论的范围，成为日后学者研究这一问题的出发点。

辛普森(Thomas Simpson，1710~1761)在 1755 年向皇家学会宣读的文章《在应用天文学中取若干观测平均值的好处》中试图证明，若以观测值的平均值估计真值，误差将比单个观测值要小，且随着观测次数的增加而减小。辛普森对一种极特殊的误差分布证明了其结论。他假定在一次天文测量中以秒来度量的误差只能取 0，±1，±2，±3，±4，±5，这 11 个值，取这些值的概率则以在 0 处最大，然后在两边按比例下降，直到 ±6 处为 0，即

$$P(X=i)=(6-|x|)r, i=\pm1,\pm2,\cdots,\pm5,$$

其中 $r=1/36$。

根据所给的分布，可算得单个误差不超过 1 秒的概率为 16/36=0.444， 不超过 2 秒的是 24/36=0. 667。为比较起见，他又计算出 6 个误差的平均值不超过 1 秒的概率是 0.725，不超过 2 秒的是 0.967。易见平均值的估计优于单个值。此可视为第一次从概率角度严格证明算术平均值的优良性。由此出发，辛普森得出了现今熟知的独立均匀分布和的密度函数公式。

拉普拉斯与辛普森的研究途径不同，他直接考虑误差理论的基本问题，即应取怎样的分布为误差分布，以及在确定误差分布后，如何根据未知量 θ 的多次测量结果 $\theta_1,\theta_2,\cdots,\theta_n$ 去估计 θ。拉普拉斯可能知道伽利略的有关结论，其给出误差分布 $f(x)$ 应满足类似的条件：

(1) $f(x)=f(-x)$。

(2) $x\to\infty, f(x)\to 0$ (因无限大误差的概率为 0)。

(3) $\int_{-\infty}^{+\infty} f(x)dx=1$ (因在任意两数值之间曲线下方的面积代表观测具有的误差在这两个值之间的概率)。

显然，有很多函数满足这三条性质，为确定其一，拉普拉斯作了如下推理：由条件(2)知，随着 x 的增加曲线 $f(x)$ 愈来愈平缓，因而其下降率 $-f'(x)$ 也应随 x 增加而下降。

设 $-f'(x)=mf(x)$， $x\geq 0, m>0$ 且为常数， 则可解得

$$f(x)=ce^{-mx},$$

其中 $c>0$ 且为常数。

由 $f(x)=f(-x)$，得

$$f(x)=ce^{mx}, x<0$$

再由

$$\int_{-\infty}^{+\infty} f(x)dx=1,$$

得

$$c=m/2,$$

于是

$$f(x)=\frac{m}{2}e^{-m|x|},-\infty<x<+\infty$$

这就是今日的拉普拉斯分布。然而，拉普拉斯很快发现，基于这个误差函数的计算是相当繁杂的，故也就不可能有多大的实际应用价值。后来，拉普拉斯得到一个更加复杂的函数表达式，也只好无功而返了。至此，拉普拉斯已感到无能为力了。今天看来，我们为拉普拉斯深感惋惜，因早在1780年他就推广了棣莫弗的结果(棣莫弗在1730年由二项分布的近似公式导出正态分布密度函数表达式)，得到了中心极限定理的较一般形式，但他未将这一成果应用到确定误差分布的问题上来。因此，科学的重大发现往往与人擦肩而过，即使是功力深厚、思维敏捷的大数学家，也是如此。

4.高斯和最小二乘法

德国慕尼黑博物馆的高斯(C.F.Gauss，1777~1855)油画像下写有：“他的思想深入数字、空间、自然的最深秘密；他测量星体的路径及地球的形状和自然力；他推动了数学的进展直到下个世纪”。的确，高斯是“能以九霄云外的高度按照某种观点掌握星空和深奥数学的天才。”由正态分布的导出可对高斯创造性思维略见一斑。

1809年，高斯发表论著《天体运动理论》。在该书末尾，他写了一节有关“数据结合”的问题，以极其简单的手法导出误差分布——正态分布，并用最小二乘法加以验证。关于最小二乘法，高斯宣称自1795年以来他一直使用这个原理。这立刻引起了勒让德的强烈反击，他提醒说科学发现的优先权只能以出版物确定， 并严斥高斯剽窃了他人的发明。他们间的争执延续了多年。因而，这两位数学家之间关于优先权的争论，在数学史上的知名度仅次于牛顿和莱布尼兹之间关于微积分发明权的争论。现在一般认为，二人各自独立地发明了最小二乘法，尽管早在10年前，高斯就使用这个原理，但第一个用文字形式发表的是勒让德。

高斯较之于勒让德把最小二乘法推进得更远， 他由误差函数推导出这个方法并详尽阐述了最小二乘法的理论依据。其推导过程如下：

设误差密度函数为$f(x)$，真值为x，n个独立测定值为$x_1,x_2,\cdots,x_n$，由于观测是相互独立的，因而这些误差出现的概率为

$$L(x)=L(x;x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i-x)$$

要找出最有希望的误差函数应使$L(x)$达极大，高斯认为$\bar{x}$就是x的估计值，并使$L(x)$取得极大值。对上式两端取对数得

$$\ln L(x)=\sum_{i=1}^{n}\ln f(x_i-x)$$

再对此式求导

$$\frac{d\ln L(x)}{dx}=-\sum_{i=1}^{n}\frac{f'(x_i-x)}{f(x_i-x)}$$

记

$$g(x)=f'(x)/f(x),$$

则有

$$\sum_{i=1}^{n} g(x_i-\bar{x})=0,$$

上式求对 x_i 的偏导数

$$\frac{\partial g}{\partial x_i}+\frac{\partial g}{\partial x_n}\frac{\partial x_n}{\partial x_i}=0,$$

而

$$\sum_{i=1}^{n} x_i-n\bar{x}=0,$$

有

$$\frac{\partial x_n}{\partial x_i}=-1(i\neq n).$$

则对任意 i 有

$$\frac{\partial g}{\partial x_i}=\frac{\partial g}{\partial x_n},$$

即

$$\frac{\partial g}{\partial x_i}=c \quad（c 为常数）,$$

可得

$$g(x)=cx+b.$$

$$\sum_{i=1}^{n} g(x_i-\bar{x})=\sum_{i=1}^{n}[c(x_i-\bar{x})+b]=c\sum_{i=1}^{n}(x_i-\bar{x})+nb=0,$$

因为

$$\sum_{i=1}^{n}(x_i-\bar{x})=0,$$

可推得

$$b=0,$$

则有

$$g(x)=f'(x)/f(x)=cx,$$

积分可得

$$f(x)=ke^{\frac{1}{2}cx^2},$$

由

$$\int_{-\infty}^{+\infty} f(x)dx=1,$$

应有 $c<0$，取 $c=-\frac{1}{\sigma^2}$，可得

$$k=\frac{1}{\sqrt{2\pi}\sigma},$$

则有

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{1}{2\sigma^2}x^2},$$

即正态分布 $N(0,\sigma^2)$。

这样可知$(x_1,x_2,\cdots,x_n)$的误差密度函数为

$$(\sqrt{2\pi}\sigma)^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-x)^2\}。$$

要使此式达到极大值，必须选取 $x_1,x_2,\cdots,x_n$ 之值而使表达式 $\sum_{i=1}^{n}(x_i-x)^2$ 达极小值。于是，可得 $x_1,x_2,\cdots,x_n$ 的最小二乘法估计。

在推证过程中，高斯有两个创新之处：

(1)他不像其前辈那样，采取贝叶斯式的推理方式，而是直接构造观测值的似然函数，即导出误差函数使其达极大估计量。

(2)高斯用逆向思维来思考这个问题，即先承认算术平均值 $\bar{x}$ 是所求的估计，即“如果在相同的环境和相等的管理下对任一个量经由多次直接观测确定，则这些观测的算术平均值是最希望要的值”。这是高斯大胆采用了人们千百年来的实际经验，实为高斯之独创性思维。这也正如他所说：“数学，要有灵感，必须接触现实世界”。

最小二乘法在 19 世纪初发明后，很快得到欧洲一些国家的天文学家和测地学家的广泛关注。据不完全统计，自 1805 年至 1864 年的 60 年间，有关最小二乘法的研究论文达 256 篇，一些百科全书包括 1837 年出版的大不列颠百科全书第 7 版，亦收入有关方法的介绍。同时，误差的分布是“正态”的，也立刻得到天文学家的关注及大量经验的支持。如贝塞尔(F.W.Bessel，1784~1846)对几百颗星球作了三组观测，并比较了按照正态规律在给定范围内的理论误差值和实际值，对比表明它们非常接近一致。拉普拉斯在 1810 年也给出了正态规律的一个新的理论推导并写入其《分析概率论》中。正态分布作为一种统计模型，在 19 世纪极为流行，一些学者甚至把 19 世纪的数理统计学称为正态分布的统治时代。在其影响下，最小二乘法也脱出测量数据意义之外而发展成为一个包罗极大，应用极其广泛的统计模型。到 20 世纪正态小样本理论充分发展后，高斯研究成果的影响更加显著。

综上可知，勒让德和高斯发现最小二乘法是从不同的角度入手的：一个是为解线性方程组，一个是寻找误差函数；一个用的是整体思维，考虑方程组的均衡性，一个用的是逆向思维，首先接受经验事实；一个是纯代数方法，一个致力于应用。相比而言，高斯不愧为数学王子，他把最小二乘法推进得更远、更深刻，这极大地推进了数理统计学的发展。

5.棣莫弗与高斯正态分布发现的文化比较

概率论中最重要的分布——正态分布是由数学家亚伯拉罕·棣莫弗和数学王子——高斯各自独立发现的。1733 年左右，棣莫弗由二项分布的逼近推导出正态分布的密度函数表达式。1809 年，高斯在推导误差分布函数时，再次发现正态分布。两位数学家在不同的数学文化背景下，从不同的角度入手，采用不同的方法，得到相同的结论，可谓殊途同归。

数学史家克莱因(M.Kline，1908~1992)曾说："数学和科学中的巨大发展，几乎总是建立在几百年中许多人所做出的一点一滴贡献的基础上。需要有一个人来走那最高最后的几步，这个人要能足够敏锐地从纷乱的猜测和说明中清理出前人有价值的说法，有足够的想象力把这些碎片重新组织起来，并且足够大胆地制定一个宏伟的计划"　。棣莫弗和高斯都是走最后这几步的大家，他们皆是集大成者。

意大利数学家帕乔利关于"点数问题"的研究、卡尔达诺在 1526 年撰写的《游戏机遇的学说》、塔尔塔利亚的著作《论数字与测量》、1654 年费马和帕斯卡的通信、惠更斯的《论赌博中的计算》等都为棣莫弗的研究奠定了基础。尤其是雅各布·伯努利的《猜度术》为棣莫弗指出了概率论的研究方向。棣莫弗在其论述中体现了与雅各布相似的目标，即通过试验来估算概率：

"如假设某事件发生与否的可能性相同，尽管在 3000 次试验中，该事件发生在 2000 次的情形缺少出现的可能，但这并不意味着它不出现；而且一旦出现，随之产生的与平等相差悬殊的比率关系也必须接受。因此，从试验中获取结论的思维应该更好。"

雅各布在《猜度术》中建立了第一个大数定律，使概率论成为了独立的数学分支，但在频率与概率接近程度上所确定的正整数 N，其精确程度令人不太满意。其后雅各布的两个侄子尼古拉斯· 伯努利(N. Bernoulli，1687~1759)和丹尼尔· 伯努利(D.Bernoulli，1700~1782)对其也进行了研究，但估计仍过于粗糙。要对之作出更精确的估计，需要计算单个二项式的系数，棣莫弗正是从这里入手取得了成功。

关于误差理论的研究也早就开始了。天文学家伽利略可能是第一个提出随机误差概念并对其有所研究的学者。他在 1632 年出版的著作《关于两个主要世界系统的对话》中提及这个问题。尽管他用"观测误差"这个名称，但所描述的性质实则为现在的随机误差分布。

辛普森(Thomas Simpson，1710~1761)在 1755 年向皇家学会宣读的文章《在应用天文学中取若干观测平均值的好处》中第一次从概率角度严格证明了算术平均值的优良性。拉普拉斯得到

$$f(x)=\frac{m}{2}e^{-m|x|},-\infty<x<+\infty$$

高斯正是在这些前辈大量工作的基础上展开研究而获得成功的。

若将

$$\lim_{n\to+\infty}P\{a\le\frac{m-n/2}{\sqrt{n}/2}\le b\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$$

变形为

$$P\{|\frac{X}{N}-\frac{1}{2}|\le\frac{c}{\sqrt{N}}\}=\frac{1}{\sqrt{2\pi}}\int_{-2c}^{2c}e^{-\frac{x^2}{2}}dx$$，

因 c 可以充分大，这就比伯努利更为简洁的证明了伯努利大数定律。

在尚无方差概念的当时，其证明的确匠心独具。此外，其精度也有所提高。如伯努利需进行 25 550 次试验的情形，而棣莫弗仅需 6498 次即可。可见棣莫弗在伯努利所开创的方向上前进了一大步。伯努利讨论的是当试验次数趋于无穷大时频率的极限行为，而棣莫弗研究的是有利事件出现的次数相关变量的极限分布。这是棣莫弗的又一个创新点。此外棣莫弗的结论还否定了过去学术上的两种极端意见，得出观察值的平均值之精度与观察次

数N的平方根($\sqrt{N}$)成正比，棣莫弗充分认识到了$\sqrt{N}$的特殊地位，故引进了“模”这个名称，后来被“标准差”取而代之。这是认识自然规律的一个重大发展。

相比之下，高斯的推证过程显得极为简洁和完善。

棣莫弗不畏艰难，历经十余载最终导出正态分布的密度函数表达式。其研究成果在概率论发展中起着承前启后的作用，尤其是在二项分布、正态分布、中心极限定理等方面，开辟了概率论发展的新方向。正是在此基础上，拉普拉斯于 1774 年，证明了

$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$，并对棣莫费的结果进行推广，建立了中心极限定理较一般的形式，即今天的棣莫费——拉普拉斯极限定理：

若随机变量$\xi \sim B(n,p)$，则对任意有限实数有a,b，$a \le b$，或$a=-\infty$，$b=+\infty$，

$$\lim_{n\to+\infty} P\{a \le \frac{\xi - np}{\sqrt{np(1-p)}} \le b\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

再有，若$0<p<1, q=1-p$，则对任意给定的实数$a<b$，当k满足$\frac{k-np}{\sqrt{npq}} \in (a,b)$时，下述极限一致地成立

$$\lim_{n\to+\infty} \frac{C_n^p p^k q^{n-k}}{(2\pi npq)^{-\frac{1}{2}} e\{-\frac{(k-np)^2}{2npq}\}} = 1$$

，

这称之为局部极限定理。

即当n很大，$|k-np|$相对于$\sqrt{npq}$不很大时，则可用

$$(2\pi npq)^{-\frac{1}{2}} e\{-\frac{(k-np)^2}{2npq}\}$$

来近似$C_n^p p^k q^{n-k}$。

利用棣莫弗所得结果可以解决一大类概率问题。如若某奖券的量很大，中奖率为 5%，如果要保证中奖的概率达 90%，至少应购买多少张奖券?类似的，也可以应用于保险业中。棣莫弗将其结果大量用于诸如此类的问题。这就初步显示了概率论这一数学分支的广泛应用性。他自己认为解决了这样的哲学问题：在人们以为纯粹偶然的事件中，可以寻找其规律和必然。正如他在《机会学说》第三版所言，尽管机会具有不规则性，由于机会无限多，随着时间的推移，不规则性与秩序性相比将显得微不足道。他说，这种秩序自然是从“固有设计中”产生出来的。

可惜的是棣莫弗的这一研究成果在相当长的时间里被遗忘了。直至 1924 年 K.皮尔逊撰《正态曲线史》一文后，人们才重新认识到他的重大贡献。

由于棣莫弗的工作被人遗忘，加之高斯所做的工作对后世影响极大，而使正态分布有了“高斯分布”之称。在此基础上，高斯导出了数理统计学中最重要的方法——最小二乘法并详尽阐明了其原理，这极大地推动了数理统计学的发展。相关回归分析、方差分析和线性模型理论等都是最小二乘法的应用或扩展。高斯所导出的正态分布为误差分布函数，立刻得到天文学家的关注及大量经验的支持。如贝塞尔对几百颗星球作了三组观测，并比较了按照正态规律在给定范围内的理论误差值和实际值，对比表明它们非常接近一

致。拉普拉斯在1810年也给出了正态规律的一个新的理论推导并写入其《分析概率论》中。正态分布作为一种统计模型，在19世纪极为流行，一些学者甚至把19世纪的数理统计学称为正态分布的统治时代。20世纪正态小样本理论发展起来之后，高斯研究成果的影响变得更加日益显著起来。

德国10马克的钞票印有高斯头像和正态概率曲线，这表明在高斯对科学的众多贡献中这是最重大的一项。

综上，两位数学大师导出正态分布一个从二项分布入手，一个由误差函数突破：一个以无穷级数为工具，一个以微积分为基础：一个历经十余载的攻关，一个轻而易举、信手拈来：一个是充满技巧的证明，一个是简洁完美的推导，一个奠定了概率论的极限理论的基础，一个推动了数理统计学的发展。纵观概率论与数理统计的今日发展，棣莫弗和高斯的开创之功实是功不可没。

4.4 多元随机变量的分布产生的背景

1.多项分布

假设事件$A_1,A_2,\cdots,A_k$是不相容的，其发生的概率分别是$p_1,p_2,\cdots,p_k$，这里$p_1+p_2+\cdots+p_k=1$，若是随机变量，分别给出在n次总试验中$A_1,A_2,\cdots,A_k$ 发生的次数$n_1,n_2,\cdots,n_k$，因$\xi_1+\xi_2+\cdots+\xi_k=n$，则

$$P(\xi_1=n_1,\xi_2=n_2,\cdots,\xi_k=n_k)=\frac{n!}{n_1!n_2!\cdots n_k!}p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}$$

是随机变量$\xi_1,\xi_2,\cdots,\xi_k$的联合概率函数，这里$n_1+n_2+\cdots+n_k=n$。

由于上式是

$$(p_1+p_2+\cdots+p_k)^n$$

的多项式展开式中的一般项，这个分布是二项分布的推广，称之为多项分布。

2.多元正态分布及其产生的背景

多维正态密度的一个特例，最早见于1776年拉格朗日的著作，他是因研究多项分布概率的极限而得出这一形式的。1812年拉普拉斯在其名著《分析概率论》中，从讨论最小二乘估计的联合极限分布，也导出多维正态密度函数。但只是作为一个数学函数的名义提出来，不具备多维随机变量的概率分布的身份。

多元正态分布以多个随机变量的联合分布的身份出现，最初也是在测量误差领域。一种说法认为这最早见于1846年布拉瓦依斯发表的一篇文章。用现代的记号，他的问题可表述如下：

设$Y_i=f_i(X_1,\cdots,X_n)$，$i=1,2,\cdots,m$是一些由“直接观测值”$X_1,\cdots,X_n$所决定的“间接观测值”，$X_1,\cdots,X_n$独立，变量X_i服从正态分布$N(a_i,\lambda_i^2)$，要决定（$Y_1,\cdots,Y_n$）的联合分布。

布拉瓦依斯只讨论了$m=2$和3的情况。方法如下：

因$e_i=X_i-a_i$是误差，其值很小，故近似地有

$$\tilde{Y}_i=Y_i-f_i(a_1,\cdots,a_n)=\sum_{j=1}^{n}A_{ij}e_j, i=1,2,\cdots$$

这里A_{ij}是$f_i(x_1,\cdots,x_n)$对x_j的偏导数在$(a_1,\cdots,a_n)$点处之值， $e_1,\cdots,e_n$独立，且e_i有分布$N(0,\lambda_i^2)$。以此为出发点，经过一些运算，他导出$(\tilde{Y}_1,\tilde{Y}_2)$的联合分布密度函数为

$$\frac{1}{\pi}\sum{}^{*}\frac{(A_{1j}A_{2k}-A_{1k}A_{2j})^2}{h_jh_k}\exp\{-\frac{\frac{y_1^2\sum_{j=1}^{n}A_{2j}^2}{h_j}-\frac{2y_1y_2\sum_{j=1}^{n}A_{1j}A_{2j}}{h_j}+\frac{y_2^2\sum_{j=1}^{n}A_{1j}^2}{h_j}}{\sum{}^{*}\frac{(A_{1j}A_{2k}-A_{1k}A_{2j})^2}{h_jh_k}}\} \tag{4—11}$$

这里$h_j=(2\lambda_j^2)^{-1}$，$\sum{}^{*}$表示对$j,k=1,2,\cdots,n,j\neq k$求和。可惜的是，布拉瓦依斯停留在这一步，没有化简到能用$\tilde{Y}_1,\tilde{Y}_2$二阶矩表达的式子。现在容易通过简单计 算证明：若以σ_1^2和σ_2^2分别记$\tilde{Y}_1,\tilde{Y}_2$的方差，ρ记其相关系数，则上式化为我们现在熟悉的形式

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}(\frac{y_1^2}{\sigma_1^2}-2\rho\frac{y_1y_2}{\sigma_1\sigma_2}+\frac{y_2^2}{\sigma_2^2})\} \tag{4-12}$$

对$m=3$， 布拉瓦依斯用类似方法得出了$(\tilde{Y}_1,\tilde{Y}_2,\tilde{Y}_3)$的联合分布，其形式类似式 4-11，但更复杂。

卡尔·皮尔逊在 1920 年写了一篇题为《相关的历史注记》的文章，其中对布拉瓦依斯的上述工作的评价不高。按皮尔逊的意见，通过间接观测值导出正态分布的作法，在高斯 1823 年出版的《数据结合理论》中已有了，且是对一般的m，而布拉瓦依斯只考虑了$m=2,5$。尤其重要的是，布拉瓦依斯未能通过变量的矩去取代式 4-11 中那些复杂的系数组合，这使他与回归相关的发现无缘。

至于多元正态作为“统计数据”（埃其渥斯的解释）的模型提出来，时间晚了很多。1885 年高尔登（在数学家狄克逊的帮助下）实际上已得出了二元正态密度的一般形式，写法与式 4-12 有所不同。对一般m元的情况，是埃其渥斯于 1892 在其题为《相关的平均值》中提出的。他把多元正态密度（期望为 0）的形式定义为

$$c\exp(-\sum_{i,j=1}^{m}a_{ij}x_ix_j) \tag{4—13}$$

其中$c>0$为常数，$A=(a_{ij})$为m阶正定方阵。他取这个形式的理由是：

假如埃其渥斯致力的问题，是要通过$(X_1,\cdots,X_m)$的 2 阶矩表达出式 4-13 中的系数a_{ij}，在$m=2$的情况这实际上已由高尔登解决了。埃其渥斯解决了$m=3$的情况并由此猜到了一般解。按我们现在习用的记号，此解可表为：

记$\sigma_{ij}=E(X_iX_j)$，$i,j=1,2,\cdots,m$（注意已假定$X_1,\cdots,X_m$有期望 0），Λ为方阵(σ_{ij})，则

$$A=\frac{1}{2}\Lambda^{-1} \tag{4—14}$$

但埃其渥斯使用了一套极为复杂的，难于理解的符号，把整个事情搞得很乱，以致他这一重要结果后来淹没了。用了一种更好的记号从而首先对式 4-14 作出清楚证明的，是卡尔· 皮尔逊（系列论文《数学用于进化论》之三，1896）。在他那个时代，矩阵表述方法还不普遍使用。皮尔逊实际是把式 4-13 写成

$$c\exp\left(-\frac{1}{2R}\sum_{i,j=1}^{m} R_{ij}\frac{x_i}{\sigma_i}\frac{x_j}{\sigma_j}\right) \tag{4—15}$$

的形式，其中 R 是相关矩阵 (ρ_{ij}) 的行列式，R_{ij} 为其代数余子式。他并给出了常数 c 的值

$$c=(2\pi)^{-\frac{m}{2}}(\sigma_1\sigma_2\cdots\sigma_m)^{-1}R^{-\frac{1}{2}} \tag{4—16}$$

而这是埃其渥斯未能做到的。

卡尔·皮尔逊把式 4-15 称为“正态相关曲面”。在他那篇很重要的有关 χ^2 拟合优度检验的论文中，他就是以这个形式为出发点。在理论上，多元正态分布的重要意义还在于：是它把起初纯属于误差分析的线性模型理论与“统计数据”的分析沟通起来。

3. χ^2 分布和 F 以及 Z 分布产生的背景

数理统计学中所谓“三大分布”之说，是指 χ^2 、t 和 F 这三个分布。此说之由来是因为它们与许多重要的统计推断问题有关。

20 世纪前 20 年，统计学的重点仍在相关回归，而这与多维正态密切联系着，而这也突出了多维正态在数理统计学中的地位。有意思的是，这三大分布的产生都与多维正态分布无关，相反，在一定意义上可以说，它们真正的根子是高斯线性模型——即在 $y=x'\beta+e$ 中视为非随机的已知向量那种线性模型。

事实上，χ^2 分布作为描述统计量的分布，最初是从线性模型最小二乘法的残差平方和分布问题导出的，比卡尔·皮尔逊的 χ^2 检验早。Student 的 t 分布可以认为是与线性模型 $y=\beta+e$ 联系着，e 服从正态分布 $N(0,\sigma^2)$。而 F 分布，则系出自高斯线性模型中变量 x 的离散化。

如果我们权且把多维正态也纳入到“线性统计模型”这个大体系内，则大致可以说，这个体系自 19 世纪初直至 20 世纪末尾，代表人物有高斯及拉普拉斯、勒让德等人，形式是误差论并逐渐渗入到统计数据分析问题。第二阶段从 19 世纪末到 20 世纪 20 年代初期，代表人物主要是卡尔·皮尔逊，形式是把多元正态与这模型联系起来（这是由于多元正态的一个特殊性质：其回归为线性且条件方差保持常数），重点转到相关回归。第三阶段可以说是自 1922 年，代表人物是 Fisher，形式是回复到以自变量为非随机并离散化，重点问题是方差分析（协方差分析）并联系到试验设计的发展，可以说，弄清了这个模型发展的脉络，也就大体上懂得了自 19 世纪初以来统计学发展的主流。

再回到 F 分布的正题。这个问题要溯源到 1917 年斯卢茨基（E.Slutsky）的一篇文章，其中提出了运用皮尔逊的拟合优度（goodness of fit）思想去检验回归是否为线性的问题。采用现在通行的记号，斯卢茨基的原假设可写为如下的模型

$$y_{ij}=x_i'\beta+e_{ij}, i=1,2,\cdots k, j=1,2,\cdots,w_i, \tag{4—17}$$

其中 $\{e_{ij}, i=1,2,\cdots k, j=1,2,\cdots,w_i\}$ 全体独立，$e_{i1},\cdots,e_{iw_i}$ 同分布且有期望 0 和方差 σ_i^2。

这里容许误差方差量取值与x_i有关，是其一特点，另一个特点是在一个自变量x_i处重复作若干次观察，其背景是：当时盛行考虑分组数据，若组范围足够小，同一组数据的值可以认为即是组的中心所在。另外，在这模型中自变量 并无随机性，这与皮尔逊学派的取法不同。

如果模型式 4-17 成立，则回归$y = x'\beta$是线性。斯卢茨基的想法是：要从数据出发去构造一个能反映与这个假设的差距的量。斯卢茨基的作法如下；

算出

$$\overline{y_i}=\frac{y_{i1}+\cdots+y_{iw_i}}{w_i},i=1,2,\cdots k$$，

及“组内方差”

$$s_k^2=\frac{\sum_{j=1}^{w_i}(y_{ij}-\overline{y_i})^2}{w_i}$$，

用加权最小二乘法

$$\sum_{i=1}^{k}w_i(y_i-x_i'\beta)^2=最小 \qquad (4—18)$$，

确定β的估计$\hat{\beta}$。计算在各点的残差$r_i = y_i - x_i'\hat{\beta}$，$i = 1,2,\cdots k$。

斯卢茨基认为，在原假设（回归为线性）成立时，统计量

$$\xi=\frac{\sum_{i=1}^{k}w_ir_i^2}{s_i^2}$$

将服从自由度为k的χ^2布。于是若就一组具体样本算出ξ之值为ξ_0。则按卡尔·皮尔逊1900年关于拟合优度检验的文章，数据与（线性）模型的拟合优度为$P(\chi_k^2 \geq \xi_0)$（当时还没有检验水平、功效一类的概念）。

斯卢茨基这个论断在数学上是不正确的，但其中包含了一种有价值的统计思想：s_i^2反映与模型取法无关的随机误差，而残差r_i，则不仅与随机误差有关，还与模型取得是否正确有关，模型与实际偏离愈大，r_i一般也会愈大，所以$\frac{r_i^2}{s_i^2}$这个量反映了以随机误差水平为标杆去衡量模型与实际的偏离程度：此量愈大，模型与数据的符合看上去愈差，这就是统计量ξ的实际背景，这个思想实际上也就是方差分析的精髓。这是一个例子，说明在评价一件统计学研究工作时，首先的要看它在统计思想和方法上有无创新。数学上的正确与否当然重要，但仍只能说是第二位的。

且说 Fisher 抓住斯卢茨基这个想法，但在数学上作了改进。首先，他假定误差服从正态分布且方差$\sigma_1^2 = \sigma_2^2$不依赖i。这样就没有必要用各个s_i^2分别除$w_ir_i^2$，而可把它们加起来，得到一个总的反映模型偏差的量

$$G_1=\sum_{i=1}^{k}w_ir_i^2$$.

同样，为估计反映随机误差水平的量σ^2，可以把各个s_i^2结合起来，因而引进

$$G_2=\sum_{i=1}^{k}w_is_i^2=\sum_{i=1}^{k}\sum_{j=1}^{w_i}(y_{ij}-\overline{y_i})^2$$

对G_2，在研究 student-t 分布中已经证明它有分布$\sigma^2\chi_{n-k}^2$，其中$n=w_1+\cdots+w_k$。

对G_1，Fisher 是这样推理的：它是k个量的平方和，本应有χ^2分布$\sigma^2\chi_k^2$。但由于$r_1,\cdots,r_k$受到p个约束(p是式 4-17 中β的维数)：事实上，式 4-18 式的加权最小二乘得出

$$\sum_{i=1}^{k}w_ix_{ij}(\overline{y_i}-x_i'\hat{\beta})=\sum_{i=1}^{k}w_ix_{ij}r_i, i=1,2,\cdots k$$，

这里$x_i'=(x_{i1},\cdots,x_{ip})$。

由此 Fisher 断言，自由度应减少p，即G_1有分布$\sigma^2\chi_{k-p}^2$。最后，Fisher 肯定G_1与G_2独立。他这一点的论据充足：G_1只与$y_1,\cdots,y_k$有关，而G_2只与$s_1^2,\cdots,s_k^2$有关。按对样本的假定，$(\overline{y_1},s_1^2),\cdots$，$(\overline{y_k},s_k^2)$独立。而在 student 的$t$分布的推导中已知$\overline{y_i},s_i^2$独立。

这样，把G_1与G_2分别除以其自由度，得到统计量

$$Q=\frac{\frac{G_1}{k-p}}{\frac{G_2}{n-k}}$$，

Fisher 指出（在原假设成立之下）它服从皮尔逊Ⅵ型分布，现今我们通称自由度为$(k-p,n-k)$的 F 分布$F_{k-p,n-k}$。但 Fisher 习惯于考虑

$$Z=\frac{1}{2}\log Q$$，

其分布在统计上成为 Z 分布。

有了这个分布就可以按皮尔逊的方式计算拟合优度：

若由数据算得$Q=Q_0$，则概率$P(F_{k-p,n-k}\geq Q_0)$愈小，数据与模型拟合愈差。

以上就是 Fisher 在 1922 年发表的论文《回归公式的拟合优度及回归系数的分布》一文中的主要内容。其所以把回归系数分布与上述内容合在一篇文章中，原因看来如下：设回归系数真值为β而其（最小二乘）估计为$\hat{\beta}$，偏差$\hat{\beta}-\beta$，它应当以随机误差为标杆去衡量。因为此处是在模型假定为正确的基础上去讨论，故随机误差方差就用残差平方和（除以自由度）去估计，取其比值即得出t分布。这个想法直接推广到检验多个回归系数的情况。要注意的是，这里的自由变量是认为非随机的。

Fisher 在文章中对G_1分布在论证没有多着笔墨，看来他基本上依靠直观看出了这个结果，即在由他所首创的“自由度”这个重要概念上。这个概念源出于他的n维几何。他在早期与 student 通信讨论$\sum_{i=1}^{n}(x_i-\overline{x})^2$的除数应是$n$或$n-1$（student 是用$n$）时，他主张用$n-1$，理由是定了$\overline{x}$后，点$(x_1,\cdots,x_n)$只能在一个通过点（$\overline{x_1},\cdots\overline{x_n}$）的$n-1$维超平面

上活动，或者说，点受到一个约束$\sum_{i=1}^{n}(x_i-\bar{x})^2=0$。因此只有$n-1$个自由度。这个分析方法他曾多次用于各种问题。实际上直到现在，当人们要确定某个二次型统计量的正确除数时，自由度的分析仍是一个便捷的方法。现在我们在讨论有关线性模型的理论问题时，经常采用化作典则形式的方法。在典则形式下，自由度通过空间维数的变化清晰可见。

4.t分布和维希特分布产生的背景

哥色特，其笔名 Student 比他的真名更为人所知。奈曼曾指出，许多统计学家在哥色特于 1937 年去世后，尚不知他就是 Student，因此我们也从众，在下文中用 Student 来称呼他。早期统计文献习惯用“Student 分布”这个称呼。用t表示 Student 的统计量，大概始于 1924 年费歇尔的文章。

哥色特 1876 年出生于坎特伯雷。他曾在温彻斯特大学和牛津大学就读。 1899 年作为一名酿酒师进入爱尔兰的都柏林一家啤酒厂工作，在那里他涉及到有关酿造过程的数据处理问题。

1906 到 1907 年他有 1 年的时间去皮尔逊那里学习和研究统计学。他着重关心的是由人为试验下所得的少量数据的统计分析问题，在当时这是一个全新的课 题，因为如前面曾指出的，当时统计学中占主导地位的卡尔·皮尔逊学派强调的是由自然观察得来的大量数据的统计处理。

这一研究的成果，就是前面曾多次讲到过的那篇使他名垂统计史册的论文《均值的或然误差》（以下简称《均》），发表于 1908 年的《生物计量》杂志上。如现在所周知的，他在文中提出了如下的结果：设$x_1,\cdots,x_n$是抽自正态分布$N(a,\sigma^2)$的随机样本，a和σ^2都未知。

记$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$，则$\frac{\sqrt{n}(\bar{x}-a)}{s}$服从自由度为$n-1$的$t$分布$t_{n-1}$。

表 4-1

z	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$\frac{\sqrt{7}}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{7x^2}{2}}dx$
.1	.5622	.5745	.5841	.5928	.6006	.60787	.61462	.60411
.2	.6241	.6458	.6634	.6798	.6936	.70705	.71846	.70159
.3	.6804	.7096	.7340	.7549	.7733	.78961	.80423	.78641
.4	.7309	.7657	.7939	.8175	.8376	.85465	.86970	.85520
.5	.7749	.8131	.8428	.8667	.8863	.90251	.91609	.90691
.6	.8125	.8518	.8813	.9040	.9218	.93600	.94732	.94375
.7	.8840	.8830	.9109	.9314	.9468	.95851	.96747	.96799
.8	.8701	.9076	.9332	.9512	.9640	.97328	.98007	.98253
.9	.8915	.9269	.9498	.9652	.9756	.98279	.98780	.99137
1.0	.9092	.9419	.9622	.9751	.9834	.98890	.99252	.99820
1.1	.9236	.9537	.9714	.9821	.9887	.99280	.99539	.99926
1.2	.9354	.9628	.9782	.9870	.9922	.99528	.99713	.99971
1.3	.9451	.9700	.9832	.9905	.9946	.99688	.99819	.99986

续表 4-1

z	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$\frac{\sqrt{7}}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{7x^2}{2}}dx$
1.4	.9531	.9756	.9870	.9930	.9962	.99791	.99885	.99989
1.5	.9598	.9800	.9899	.9948	.9973	.99859	.99926	.99999
1.6	.9653	.9836	.9920	.9961	.9981	.99903	.99951	
1.7	.9699	.9864	.9937	.9970	.9986	.99933	.99968	
1.8	.9737	.9886	.9950	.9977	.9990	.99953	.99978	
1.9	.9770	.9904	.9959	.9983	.9992	.99967	.99985	
2.0	.9797	.9919	.9967	.9986	.9994	.99976	.99990	
2.1	.9821	.9913	.9973	.9989	.9996	.99983	.99993	
2.2	.9841	.9941	.9978	.9992	.9997	.99987	.99995	
2.3	.9858	.9950	.9982	.9993	.9998	.99991	.99996	
2.4	.9873	.9957	.9985	.9995	.9998	.99993	.99997	
2.5	.9886	.9963	.9987	.9996	.9998	.99995	.99998	
2.6	.9898	.9967	.9989	.9996	.9999	.99996	.99999	
2.7	.9908	.9972	.9991	.9997	.9999	.99997	.99999	
2.8	.9916	.9975	.9992	.9998	.9999	.99998	.99999	
2.9	.9924	.9978	.9993	.9998	.9999	.99998	.99999	
3.0	.9931	.9981	.9994	.9998	–	.99999	–	

《均》文一开始有一段很长的导言，说明他考虑这个问题的动因，大略是：众所周知，当样本量很大时，基于正态（即认为 $\frac{\bar{x}}{s}$ 为正态分布——本书作者注）的方法是可信的，但没有人很清楚地告诉过我们：样本量的“大”和“小”的界限在哪里，而本文的目的是定出这样的一个界限。正文的主要内容有：推导出 $\frac{\bar{x}}{s}$ 的分布；计算出其标准差为 $\frac{1}{\sqrt{n-3}}$ 及峰度系数为 $3+\frac{2}{n-5}$（应为 $3+\frac{6}{n-5}$，又 t_{n-1} 的标准差为 $\sqrt{\frac{n-1}{n-3}}$。我们要记住 Student 讨论的 $\frac{\bar{x}}{s}$ 与 t_{n-1} 的差别）；计算了一个小型的（$\frac{\bar{x}}{s}$ 的）分布表如表 4—1 所示，最后给了几个实用例子。

表 4—1 的用法是（要记住已假定总体均值为 1）：

例如，设 $n=7$，则 $P(\frac{\bar{x}}{s}\leqslant 0.6)=0.9\,040$。最末一列就是 $n=7$ 的情况，对 $\frac{\bar{x}}{s}$ 的正确分布与其近似的正态分布的比较。如就上例，$n=7$ 时，$P(\frac{\bar{x}}{s}\leqslant 0.6)$的正确值为 0.9040。但若近似地认为 $\frac{\bar{x}}{s}\sim N(0,7)$，则这个概率将是 0.94 375。

表 4-1 的历史意义在于，它是应用上极其重要的 t 分布的第一张表。后来在 1917 年 Student 又对表进行了少许扩充。限于当时的计算条件及 t 密度积分计算 的复杂性，表中

的结果略有误差。Student 自己在 1923 年核验了这两个表，结论 说“二者完全不行”(“both perfectly rotten”)。 按上表来比较几个值，算 $P(t_q \le t)$(相当于$n=10$)：

(1)$a=0.3$：0.61 462， 正确值 0.61 451。

(2)$a=0.6$：0.71 846， 正确值 0.71 835。

(3)$a=0.9$：0.80 423， 正确值 0.80 422。

其他值的比较相当，看出误差只在10^{-4}，对于应用毫无影响。今天我们面对这张表，考虑到他当年简陋的计算条件且 Student 本人并非数学出身，能算出有这种精度的结果，可以设想他付出了多少经历及其工作态度之认真，因而不由得要表示赞赏。

现在我们来讨论一下《均》文的核心部分，即 Student 是如何得出他的分布的。 他的证明分为三步：

(1)找s^2的分布。做法是：先算出s^2的偏度系数$\beta_1=\dfrac{8}{n-1}$，峰度系数$\beta_2=\dfrac{3(n+3)}{n-1}$，得到

$$2\beta_2-3\beta_1-6=0,$$

据此，他推断：“s^2的分布可望拟合一个属于皮尔逊Ⅲ型的分布。”按矩法定义出s^2的密度为

$$cx^{\frac{n-3}{2}}e^{\frac{-nx}{2\sigma^2}}(c>0\text{为常数})。$$

(2)证明$\bar{x}^2$与s^2不相关。 这通过计算相关系数容易得出。

(3)据(2)，用独立变量(显然，在当时 Student 必定还不明白“不相关”和“独立”不是一回事，虽然在这里二者碰巧是一致的。 对这一点不能苛责他。大家如卡尔·皮尔逊，甚至到 1920 年代对此尚未明确。《奈曼——现代统计学家》一书第 83 页讲了一段与此有关的故事。商的密度公式算$Z=\dfrac{\bar{x}}{s}$的密度，由于$\bar{x}$、s的密度都已知悉，这个计算不难。

如今，粗通概率统计的人也能指出 Student 推导中的漏洞之所在。 最早注意到这个问题的是 Fisher，他于 1912 年向他的天文学家老师谈到这个问题。 后者正好认识 Student，因而建议 Fisher 直接与 Student 联系，这样就开始了两人的通信及长达二十余年的友谊。Student 开始对 Fisher 的论点有所忧疑并曾为此事写信给卡尔·皮尔逊商量。在 Fisher 致 Student 的第 3 封信中他给出了完整的证明并显示已使 Student 相信。不过，Fisher 的证明迟至 1925 年才正式发表。

这个插曲的一个重大的历史后果是，Fisher 因此发展了其“n维几何”的方法，他发现这在正态样本统计量的抽样分布中，是一个极其有力的方法。 沿用这个方法，Fisher 获得了一些在应用上极重要的统计量的精确分布，它促成了统计学的“MarkII”阶段的加速到来，其意义十分重大。

Fisher 的“n维几何”法就是把样本$(x_1,\cdots,x_n)$看成n维欧氏空间R^n中的一点。 这点落在一个元区域内的概率就是分布的概率元。如在本例，要求$\bar{x}$、s的联合分布(仍设总体均值为 0)，则要设法找出在R^n中，集合

$$\{(x_1,\cdots,x_n)|\xi_0 \le \xi \le \xi_0+\Delta\xi_0, \eta_0 \le \eta \le \eta_0+\Delta\eta_0\} \quad (4-19)$$

是怎样一个形状，这里

$$\xi=\sqrt{n}\bar{x}, \eta=\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

在 R^n 中过原点与 ξ_0 作一条射线 OB ，ξ_0 为点 $(\bar{x},\cdots,\bar{x})$ 。图 4-1 表示 $n=3$ 的情况，图中的 M 是 ξ_0 ，而 P 是样本点 $(x_1,\cdots,x_n)$。过 M 点作 $n-1$ 维超平面与 OM 垂直，则 P 点位于此平面上以 M 为中心，以 $\eta_0=MP$ 为半径的超球面上，超球面的维数 $n-2$ (在 $n=3$ 时，此超球面为图中的圆周，是 1= 3 － 2 维)。

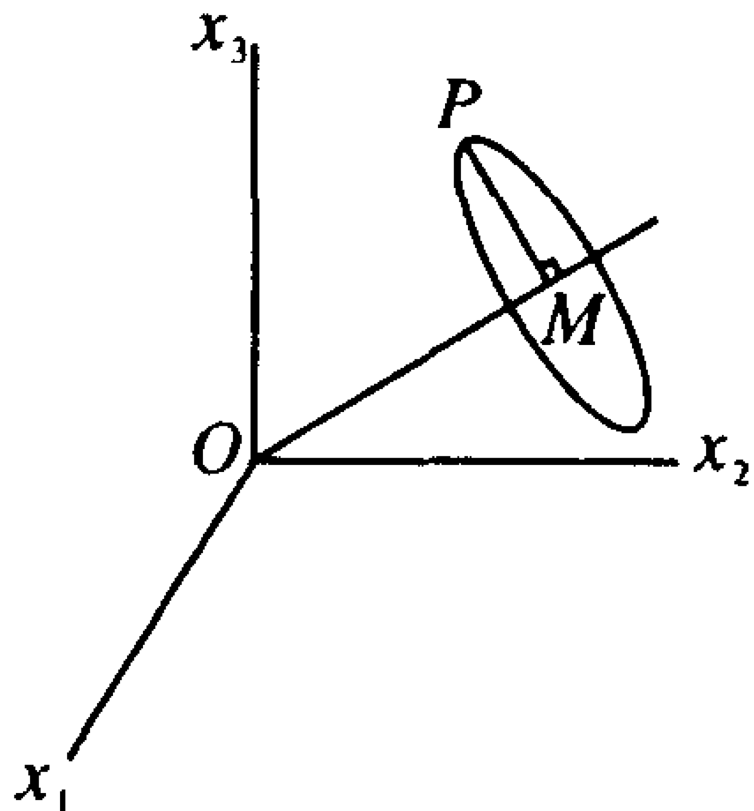


图 4-1

现如 η 在 η_0 到 $\eta_0+\Delta\eta_0$ 内变化，则其区域相当于两球面之间的体积元，其体积为 $c\eta_0^{n-2}\Delta\eta_0$。

另外 ξ 在 ξ_0 到 $\xi_0+\Delta\xi_0$ 内变化，等于说这个向度上还有一个 $\Delta\xi_0$ 的变化幅度。 由于 OM 轴与 P 所在的超平面正交，知集合式 4-19 的体积元为

$$c\eta_0^{n-2}\Delta\xi_0\Delta\eta_0,$$

而样本密度在上述体积元内（基本上）是一常数， 此因

$$c\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2)=c\exp\{-\frac{1}{2\sigma^2}(n\bar{x}^2+\sum_{i=1}^{n}(x_i-\bar{x})^2)\}$$

$$=c\exp(-\frac{1}{2\sigma^2}\xi_0^2)\exp(-\frac{1}{2\sigma^2}\eta_0^2),$$

这与上述体积元表达式结合，得出集合式 4-19 的概率元有表达式

$$=c\exp(-\frac{1}{2\sigma^2}\xi_0^2)\exp(-\frac{1}{2\sigma^2}\eta_0^2),$$

这一举证明了 $\sqrt{n}\bar{x}$ 与 s 独立，前者有正态分布 $N(0,\sigma^2)$ 而后者有皮尔逊Ⅲ型分布 $cs^{n-2}e^{-\frac{ns^2}{2\sigma^2}}$ (仍以 $\frac{\eta}{\sqrt{n}}$ 作为 s 的定义)，与 Student 猜到的完全一致。

这个例子描述的 Fisher 的“n 维几何法”，适用于其他更复杂的情况。当然，在其他情况，体积元的寻求复杂得多。 Fisher 从小训练出来的几何直观帮了他很大的忙，例如对相关系数 r 的分布，其分析推导极其复杂。Fisher 借助几何直观大大简化了推理过程。这个方法的另一个突出的应用例子，是 1928 年维希特基于此法导出了任意维正态样本全体二阶矩的联合分布——维希特分布。

重要的是 Student 所提问题的实际背景，即他首次把小样本问题提到日程上。然而，随着小样本理论的进展，其重要意义日益为统计学界所理解，特别是 t 分布的意义，因为这个分布以后多次出现在一些重要统计量分布的结果中，于是 Student 这一结果的行情逐日看涨，导致后来统计界将他尊为小样本理论的开创者和鼻祖。从 Student 的工作意义和对以后数理统计学发展所起的影响来看，应该说对他这一评价是当之无愧的。

第 5 章　极限理论

概率论极限理论是概率论的主要分支之一，也是概率论的其它分支和数理统计的重要基础。前苏联著名概率论学者 Gnedenko 和 Kolmogrov 曾说过：“概率论的认识论的价值只有通过极限定理才能被揭示，没有极限定理就不可能去理解概率论的基本概念的真正含义。”

5.1　概率论中三个重要分布的关系

1.定义

（1）设ξ是任一随机变量，称$\varphi(t)=Ee^{it}\xi$，$-\infty<t<+\infty$是ξ的特征函数。

（2）若随机变量ξ的特征函数为

$$\varphi(t)=(pe^{it}+q)^n, p、q>0, p+q=1 \text{，} \tag{5-1}$$

则称ξ服从二项分布。

（3）若随机变量ξ的特征函数为

$$\varphi(t)=e^{\lambda}(e^{it}-1), \lambda>0 \text{，} \tag{5-2}$$

则称ξ服从泊松分布。

（4）若随机变量ξ的特征函数为

$$\varphi(t)=e^{-\frac{t^2}{2}} \text{，} \tag{5-3}$$

则称ξ服从正态分布。

2.定理与证明

定理 1（泊松定理）：在n重伯努利试验中，事件A在一次试验中出现的概率为p（与试验总数n有关）。如果$n\to+\infty$时，$np_n\to\lambda$（$\lambda>0$是常数），则有

$$\lim_{n\to+\infty} b(k;\ np_n)=\frac{\lambda^k}{k!}e^{-\lambda} \tag{5-4}$$

证明：记$np_n=\lambda_n$，因为二项分布的特征函数为

$$\begin{aligned}\varphi_n(t)&=(p_ne^{it}+q_n)^n\\&=[1+\frac{np_n(e^{it}-1)}{n}]^n\\&=[1+\frac{\lambda_n(e^{it}-1)}{n}]^n\text{，}\end{aligned}$$

所以

$$\lim_{n\to+\infty}\varphi_n(t)=\lim_{n\to+\infty}[1+\frac{\lambda_n(e^{it}-1)}{n}]^n\text{，}$$

故由 Levy 定理知

$$\lim_{n\to+\infty} b(k,\ np_n)=\frac{\lambda^k}{k!}e^{-\lambda}$$

成立。

定理 2（棣莫弗—拉普拉斯定理）：在 n 重伯努利试验中，事件 A 在一次试验中出现的概率为 p（$0<p<1$），μ_n 为 n 次试验中事件 A 出现的次数，则有

$$\lim_{n\to+\infty} P(\frac{\mu_n-np}{\sqrt{npq}}<x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt \quad (5\text{-}5)$$

证明：用 ξ_k 表示事件 A 在 k 次试验中出现的次数，则

$$\xi_k=\begin{cases}1, A出现\\0, A不出现\end{cases},$$

于是

$$E\xi_k=p, D\xi_k=pq(k=1,2,\cdots).$$

设 ξ_k-p 的特征函数是 $\varphi(t)$，则

$$\frac{\sum_{k=1}^{n}\xi_k-np}{\sqrt{npq}}=\sum_{k=1}^{n}\frac{\xi_k-p}{\sqrt{npq}},$$

的特征函数为

$$[\varphi(\frac{t}{\sqrt{npq}})]^n.$$

又因为

$$E(\xi_k-p)=0, D(\xi_k-p)=pq, k-1,2\cdots,$$

所以

$$\varphi'(0)=0, \varphi''(0)=-pq$$

于是特征函数 $\varphi(t)$ 有展开式

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\varphi''(0)\frac{t^2}{2}+o(t^2)$$
$$=1-\frac{1}{2}pqt^2+o(t^2),$$

从而对任意固定的 t，有

$$[\varphi(\frac{t}{\sqrt{npq}})]^n=1-\frac{t^2}{2n}+o(\frac{t^2}{n})\to e^{-t^2/2}, n\to+\infty.$$

故由 Levy 定理知

$$\lim_{n\to+\infty} P(\frac{\mu_n-np}{\sqrt{npq}}<x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

成立。

定理 3：若 ξ_k 是服从次数为 λ 的泊松分布的随机变量，则

$$\lim_{n\to+\infty} P(\frac{\xi_k-\lambda}{\sqrt{\lambda}}<x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt \quad (5\text{-}6)$$

证明：已知ξ_k的特征函数为

$$\varphi(t)=e^{\lambda(e^{it}-1)},$$

故$\dfrac{\xi_k-\lambda}{\sqrt{\lambda}}$的特征函数为

$$g_{\lambda}(t)=e^{\lambda(e^{\frac{it}{\sqrt{\lambda}}}-1)-i\sqrt{\lambda}t},$$

对任意的t，有

$$e^{\frac{it}{\sqrt{\lambda}}}=1+\frac{it}{\sqrt{\lambda}}+\frac{t^2}{2!\lambda}+o(\frac{1}{\lambda}),\lambda\to\infty,$$

于是

$$\lambda(e^{\frac{it}{\sqrt{\lambda}}}-1)-i\sqrt{\lambda}t=-\frac{t^2}{2}+\lambda o(\frac{1}{\lambda})\to-\frac{t^2}{2}\lambda\to\infty,$$

从而对于任意的点列$\lambda_n\to\infty$，有

$$\lim g_{\lambda_n}(t)=e^{-\frac{t^2}{2}}.$$

因为λ_n是可以任意选取的，由 Levy 定理知

$$\lim_{n\to+\infty}P(\frac{\xi_k-\lambda}{\sqrt{\lambda}}<x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

成立。

3.结论

定理 1 表明泊松分布可用二项分布来逼近，定理 2 表明二项分布收敛于正态分布，而定理 3 则表明泊松分布也收敛于正态分布。总之，三个定理揭示了二项分布、泊松分布及正态分布之间的关系，同时表明在这三个重要分布中，正态分布占有头等重要的地位。其相互关系可用图 5-1 简单表示出来。

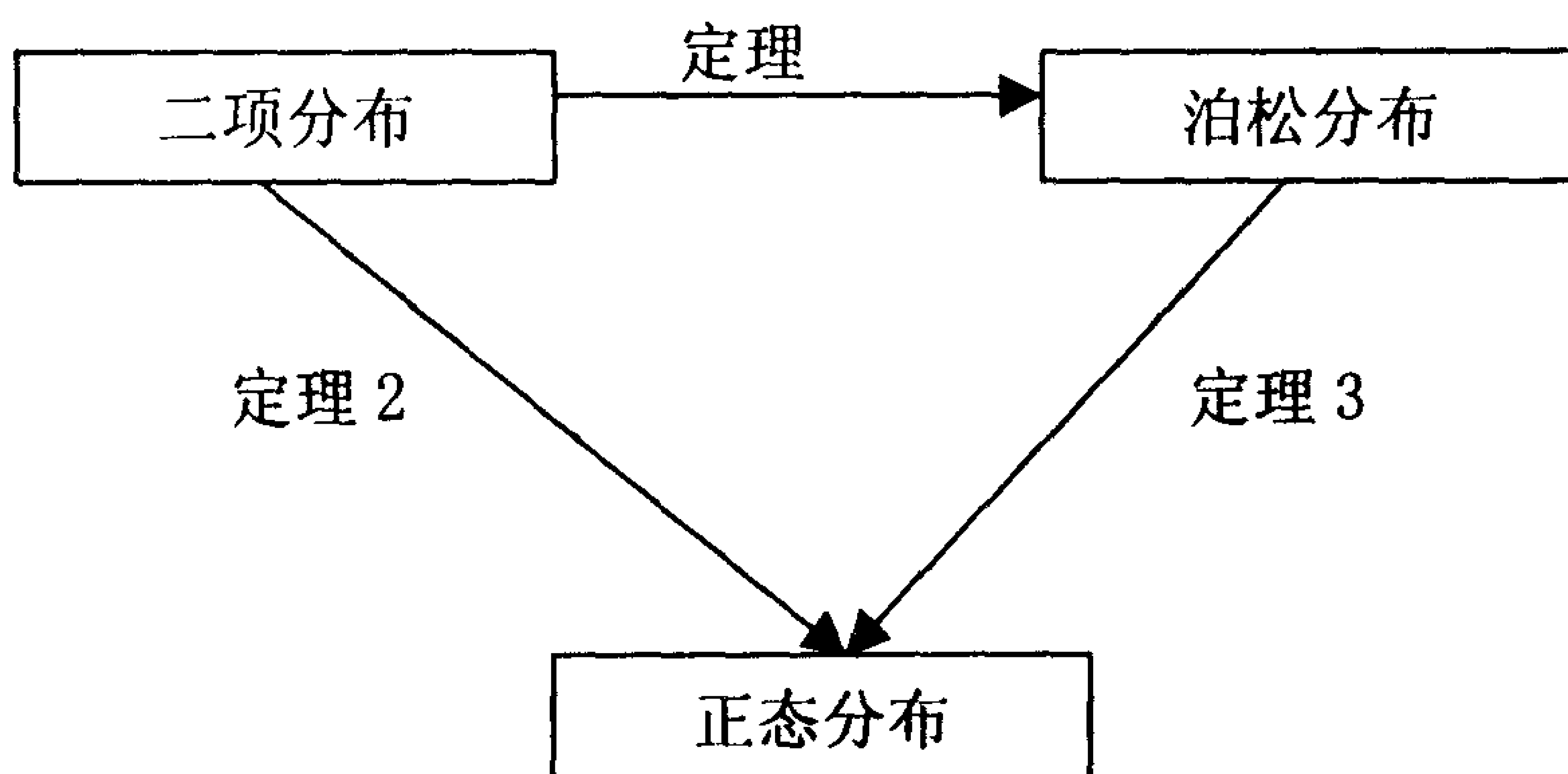


图 5-1

5.2 中心极限定理

在客观实际中有许多随机变量，它们是由大量的相互独立的随机因素的综合影响而形成的。而其中每一个别因素在总的影响中所起的作用都是微小的。这种随机变量往往近似地服从正态分布。这种现象就是中心极限定理的客观背景。在数理统计中我们看到，中心极限定理是大样本统计推断的理论基础。

1.两个实例

（1）先以重复扔对称硬币为例。重复扔 10 次，出现 0 次，1 次，……，乃至 10 次正面的概率容易计算如下

$$p_k=\frac{C_{10}^k}{2^{10}}(k=0,1,\cdots,10)$$

我们得到表 5-1，把它们画成直方图 5-2，这个图以 5.5 为对称轴。

表 5-1

p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
0.001	0.01	0.045	0.12	0.21	0.25	0.21	0.12	0.045	0.01	0.001

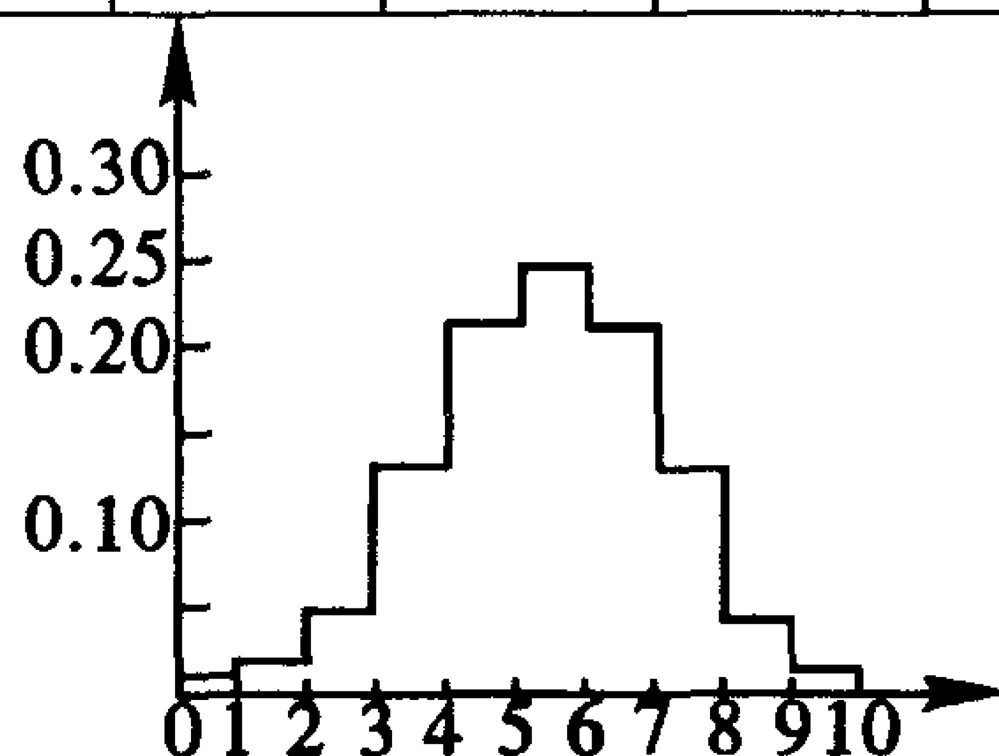


图 5-2

如果我们把重复扔的次数增加到 10000 次再画直方图，横坐标有 10001 段，而对称轴在 5000.5 处，最高直方约为

$$\frac{1}{100\sqrt{\pi}}\approx 0.0056$$

但当我们把直方图所有高度放大$\frac{100}{\sqrt{2}}$倍，同时把横底缩小$\frac{100}{\sqrt{2}}$倍，于是所有直方的顶端几乎可以联成一条连续曲线，这条曲线的方程是下列曲线的平移

$$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

此曲线是$\mu=0,\sigma^2=1$的正态曲线，叫做标准正态曲线。

用数学语言来表述以上的现象。不妨考虑非对称的硬币，第i次扔硬币出现的结果用随机变量X_i表示，$X_i=1$表示出现正面，$X_i=0$表示出现反面，且

$$P(X_i=1)=p(\neq\frac{1}{2}),P(X_i=0)=1-p=q$$

令 $S_n=\sum_{i=1}^{n}X_i$，它表示扔 n 次出现正面的次数。易知期望与方差为

$$EX_i=p,D(X_i)=E(X_i-EX_i)^2=pq.$$

由独立性知

$$ES_n=np,D(S_i)=E(S_n-ES_n)^2=npq.$$

用概率计算法则和斯特林（J.Stirling）公式，可以推出著名的棣莫弗–拉普拉斯中心极限定理：

对任意常数 a,b（$-\infty<a<b<+\infty$），有

$$\lim_{n\to+\infty}P\{a<\frac{S_n-np}{\sqrt{npq}}\le b\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx,$$

这也可表示为：当 $\Delta x>0$ 很小时，有

$$P\{x\le\frac{S_n-np}{\sqrt{npq}}\le x+\Delta x\}=\frac{1}{\sqrt{2\pi}}\int_x^{x+\Delta x}e^{-\frac{u^2}{2}}du$$

$$\approx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Delta x。\quad（5-7）$$

当 $p=q=\frac{1}{2}$，$n=10000$ 时，式 5-7 化为

$$P\{5000+50x\le S_n\le 5000+50(x+\Delta x)\}$$
$$=F(5000+50(x+\Delta x))-F(5000+50x)$$
$$\approx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Delta x,$$

这个近似，恰描绘了上面所说的直方图可用正态曲线逼近。

（2）如果设想硬币出现正面的概率 $P(X_i=1)=p$ 随着试验的次数 n 变化，这时把 p 记为 p_n，于是当 n 不同时，有

$$P\{S_n=k\}=C_n^k p_n^k(1-p_n)^{n-k},(k=0,1,2,\cdots,n).$$

先考虑 $p_n=\frac{\alpha}{n}$（$\alpha>0$）的情形，这时

$$P\{S_n=k\}=C_n^k(\frac{\alpha}{n})^k(1-\frac{\alpha}{n})^{n-k},(k=0,1,2,\cdots,n),$$

由于

$$\lim_{n\to+\infty}(1-\frac{\alpha}{n})^n=e^{-\alpha},$$

于是得

$$\lim_{n\to+\infty}P\{S_n=0\}=\lim_{n\to+\infty}C_n^0(\frac{\alpha}{n})^0(1-\frac{\alpha}{n})^n=e^{-\alpha}。$$

但

$$\lim_{n\to+\infty}\frac{P\{S_n=k+1\}}{P\{S_n=k\}}=\lim_{n\to+\infty}\frac{C_n^{k+1}(\frac{\alpha}{n})^{k+1}(1-\frac{\alpha}{n})^{n-k-1}}{C_n^k(\frac{\alpha}{n})^k(1-\frac{\alpha}{n})^{n-k}}=\frac{\alpha}{k+1},\qquad(5\text{-}8)$$

故由式 5-8 递推，得

$$\lim_{n\to+\infty}P\{S_n=1\}=\alpha\lim_{n\to+\infty}P\{S_n=0\}=\alpha e^{-\alpha},$$

$$\lim_{n\to+\infty}P\{S_n=2\}=\frac{\alpha}{2}\lim_{n\to+\infty}P\{S_n=1\}=\frac{\alpha^2}{2}e^{-\alpha},$$

……

$$\lim_{n\to+\infty}P\{S_n=k\}=\frac{\alpha}{k}\lim_{n\to+\infty}P\{S_n=k-1\}=\frac{\alpha^k}{k!}e^{-\alpha},$$

……

极限分布为

$$\frac{\alpha^k}{k!}e^{-\alpha}\,(k=0,1,2,\cdots,),$$

这是以α为参数的泊松分布，若随机变量Y的分布为：

$$P\{Y=k\}=\frac{\alpha^k}{k!}e^{-\alpha}\,(k=0,1,2,\cdots,),$$

容易计算出$EY=\alpha$， $D(Y)=\alpha$，记为$Y\sim\pi$（α）。

上面推导中假设了$p_n=\dfrac{\alpha}{n}$，事实上若把此假设放松为

$$\lim_{n\to+\infty}np_n=\alpha,$$

极限分布仍是泊松分布。

二项分布、泊松分布、正态分布是常用到的三个重要分布，在上节中我们已经用特征函数法推导了三者的关系：表明泊松分布可用二项分布来逼近，二项分布收敛于正态分布，而泊松分布也收敛于正态分布。

2.中心极限定理

以上讨论了重复扔硬币的两种情形，由于对X_i的分布的假设不同，S_n的极限分布也不同。根据这两个特殊例子，可提出一般的问题：

设X_i $(i=0,1,2,\cdots,)$为相互独立随机变量序列，$S_n=\sum_{i=1}^{n}X_i$，当$n\to+\infty$时，S_n的分布如何？这个问题称为“中心极限问题”，它是概率论中最重要的一类问题，其中的定理有广泛的理论与实际意义。

把棣莫弗-拉普拉斯定理中的条件$p_n=p$取消，且对每个X_i不必限制为相同的二值分布，只假定$EX_i=\alpha_i$， $D(X_i)=\sigma_i^2$ $(i=0,1,2,\cdots,)$。这时令$s_n^2=\sum_{i=1}^{n}\sigma_i^2$，考虑$S_n$的标准化

$$S_n^* = \frac{\sum_{i=1}^{n}(X_i - \alpha_i)}{s_n}$$

（当 $p_n = p$ 时，$\alpha_i = p$，$D(X_i) = \sigma_i^2 = pq$，$s_n^2 = \sum_{i=1}^{n}\sigma_i^2 = npq$，$S_n^* = \frac{S_n - np}{\sqrt{npq}}$）。

李雅普诺夫改进了切比雪夫的矩法，用之给出了如下的定理：

对一切 a,b（$-\infty < a < b < +\infty$），若对某个正数 $\delta > 0$，能成立 $\lim_{n\to+\infty}\frac{1}{s_n^{2+\delta}}\sum_{i=1}^{n}E|X_i - \alpha_i|^{2+\delta} = 0$，则有

$$\lim_{n\to+\infty}P\{a \le S_n^* \le b\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$$，

此性质简称渐近正态性。

随着特征函数方法的引入，中心极限定理的研究得到了很快的发展。20 世纪 20 年代，林德贝格和莱维解决了诸随机变量 X_i $(k = 0,1,2,\cdots,)$ 为相同分布时的中心极限定理。设 $EX_i = \alpha$，$D(X_i) = \sigma^2$ $(k = 0,1,2,\cdots,)$，则部分和 S_n 的标准化 $S_n^* = \frac{S_n - n\alpha}{\sigma\sqrt{n}}$ 有渐近性。1935 年，林德贝格和费勒又进一步对非相同分布的独立随机变量序列进行研究，他们得到了最一般的中心极限定理，使前人所得的结果都是它的推论。

此结果叙述如下：设 $EX_i = \alpha_i$，$D(X_i) = \sigma_i^2$ $(k = 0,1,2,\cdots,)$，则部分和 S_n 的标准化 $S_n^* = \frac{S_n - n\alpha}{\sigma\sqrt{n}}$ 具有渐近正态且费勒条件

$$\lim_{n\to+\infty}\frac{\max_{1\le i\le n}\sigma_i^2}{s_n^2} = 0$$

成立的充分必要条件使如下的林德贝格条件成立：对于任意 $\tau > 0$，

$$\lim_{n\to+\infty}\frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x-\alpha_i|\ge\tau s_n}(x-\alpha_i)^2 dF_i(x) = 0$$，

其中 $F_i(x) = P(X_i \le x)$ 为随机变量 X_i $(k = 0,1,2,\cdots,)$ 的分布函数。

3.中心极限定理的发展

此后中心极限定理的研究，是围绕以下几个方面进行的；减弱对随机变量序列独立性的要求，而考虑具有某种相依性的随机变量序列；讨论向标准正态密度函数收敛的问题；向正态分布收敛的速度及有关问题;大偏差理论等问题。

（1）相依随机变量的中心极限定理。

这一问题仍是很活跃的研究方向，其中讨论较多且获得应用的有 m 相依随机变量序列，强平稳随机变量序列，鞅序列，马尔可夫过程及其泛函以及各种类型的统计量序列。在相应的充分条件下，可以证明中心极限定理成立。

（2）局部极限定理。

向正态密度函数收敛的问题很早以前就有人讨论，但直至 20 世纪中期才得到一般性

的结果。在棣莫弗-拉普拉斯定理的形成过程中，首先解决的是重复 n 次扔硬币时，正面出现次数 $S_n=k$ 的概率渐近于正态密度的问题，即在任意给定的有穷区间 (a,b) 中，对于满足 $a\le\frac{S_n-np}{\sqrt{npq}}\le b$ 的 S_n，一致地有

$$\lim_{n\to+\infty}\frac{P(S_n=k)}{(npq)^{-\frac{1}{2}}\varphi(x_k)}=1,$$

其中 $x_k=\frac{k-np}{\sqrt{npq}}$，$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$。

这个结果称为棣莫弗-拉普拉斯局部极限定理。

此问题的推广时讨论取值为 $b+Nk$ $(N=0,\pm1,\pm2,\cdots)$ 的独立随机变量序列 X_i 相应的问题，即格点极限问题。对独立同分布情形，1948 年格涅坚科给出了充分必要条件；对独立非同分布情形，于 20 世纪 50 年代也给出了充分条件，当独立随机变量序列 $\{X_i\}$ 的标准化和 S_n^* 有密度函数 $p_n(x)$ 时，讨论 $p_n(x)$ 向标准正态密度函数收敛的问题称为局部极限定理。

格涅坚科于 1953 年对独立同分布的情形，给出了十分简洁的充分必要条件，即：当且仅当存在某个正整数 N，使 $p_N(x)$ 有界时，局部极限定理成立，即

$$\lim_{n\to+\infty}|p_n(x)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}|=0.$$

对非同分布的情形，彼得洛夫给出了局部极限定理成立的充分必要条件。

（3）收敛速度问题。

20 世纪 40 年代，贝瑞和爱森先后研究了独立随机变量序列 $\{X_i\}$ 的标准化和 S_n^* 的分布函数 $F_n(x)$ 向 $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$ 趋近的速度问题，他们的结果是：

当 $EX_i=0$，$E(X_i^2)<+\infty$，$E|X_i|^3<+\infty$ $(i=0,1,2,\cdots,)$，$s_n^2=\sum_{i=1}^{n}EX_i^2$ 时，则有

$$\sup_x|F_n(x)-\Phi(x)|\le AL_n, \tag{5-9}$$

其中 $L_n=s_n^{-\frac{3}{2}}\sum_{i=1}^{n}E|X_i|^3$，$A$ 为常数。

这一不等式 5-9 给出了 $F_n(x)$ 向 $\Phi(x)$ 趋近的速度。对于这方面的研究，已进行得相当深入。

（4）大偏差定理。

对独立同分布随机变量序列 $\{X_i\}$，若 $EX_i=0$，$E(X_i^2)=\sigma^2<+\infty$，则对常数 $M>0$，有

$$\lim_{n\to+\infty}\sup_{0\le X\le M}\frac{P(S_n^*>x)}{1-\Phi(x)}=1,\lim_{n\to+\infty}\sup_{0\le X\le M}\frac{P(S_n^*\le -x)}{\Phi(-x)}=1.$$

如果上述 M 不是常数，而是随着 n 而增大的 M_n 且 $M_n \to +\infty$（当 $n \to +\infty$ 时），则与上述相类似的结果称为大偏差定理，确切地说，如果

$$\lim_{n\to+\infty}\sup_{0\le X\le M}\frac{P\left(S_n^* > x\right)}{1-\Phi(x)}=1,\lim_{n\to+\infty}\sup_{0\le X\le M}\frac{P\left(S_n^* \le -x\right)}{\Phi(-x)}=1 \quad (5\text{-}10)$$

则称对 M_n 的大偏差定理成立。

1938 年，克拉美在渐近展开的基础上证明：若存在正常数 H，使得当 $|t|< H$ 时，$Ee^{tX_i} < +\infty$，则对

$$\lim_{n\to+\infty}\frac{M_n}{\sqrt{n}}=0$$

的大偏差定理成立。

其后，林尼克给出了对 $M_n = bn^{\alpha-\frac{1}{2}}$ 的大偏差定理成立的充分必要条件，这里 b 为正常数，$\frac{1}{2} < \alpha < 1$。

大偏差定理在理论上很重要，而且直到现在还被了工作者所重视，尚在不断地向深度发展。

5.3 大数定律

事件发生的频率具有稳定性，即随着试验次数的增加，事件发生的频率逐渐稳定于某个常数。在实践中人民还认识到大量测量值的算术平均值也具有稳定性。这种稳定性就是大数定律的客观背景。

大数定律是一组定律，它的中心问题是研究随机变量序列 $\{X_n\}$ 满足什么条件时，才能服从大数定律。大数定律有“强”、“弱”之分，它们分别对应着“以概率 1 收敛”与“依概率收敛”两个概念：

（1）若 $P(\lim\limits_{n\to+\infty} X_n = X) = 1$，则称随机变量序列 $\{X_n\}$ 以概率 1 收敛于 X。

（2）若 $\lim\limits_{n\to+\infty} P(|X_n - X| < \varepsilon) = 1$，则称随机变量序列 $\{X_n\}$ 依概率收敛于 X。

对照极限的“$\varepsilon - \delta$ 说法”，可见两者的差异仅在于极限与概率交换次序而已。

以“以概率 1 收敛”为前提，可推出“依概率收敛”；从“依概率收敛”出发，又可推出“依分布收敛”。

设 $F_n(x)$、$F(x)$ 分别为随机变量 X_n、X 的分布函数，若对 $F(x)$ 的每一个连续点 x 都有

$$\lim_{n\to+\infty} F_n(x)=F(x),$$

则称 X_n 依分布收敛于 X，而中心极限定理，就是讨论随机变量序列的标准化部分和依分布收敛于正态随机变量的定理。因而大数定律与中心极限定理各自为一种普适条件，分别构成一组定律和定理。粗略地说，大数定律研究的是级数 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 的取值趋势，中心极限

定理研讨的是级数$\frac{1}{n}\sum_{i=1}^{n}X_i$的取值精度。大数定律与中心极限定理的内涵与外延如图 5-3，5-4 所示，图中的箭头方向表示约束条件越来越强，由下向上构成一种属或推论的关系。

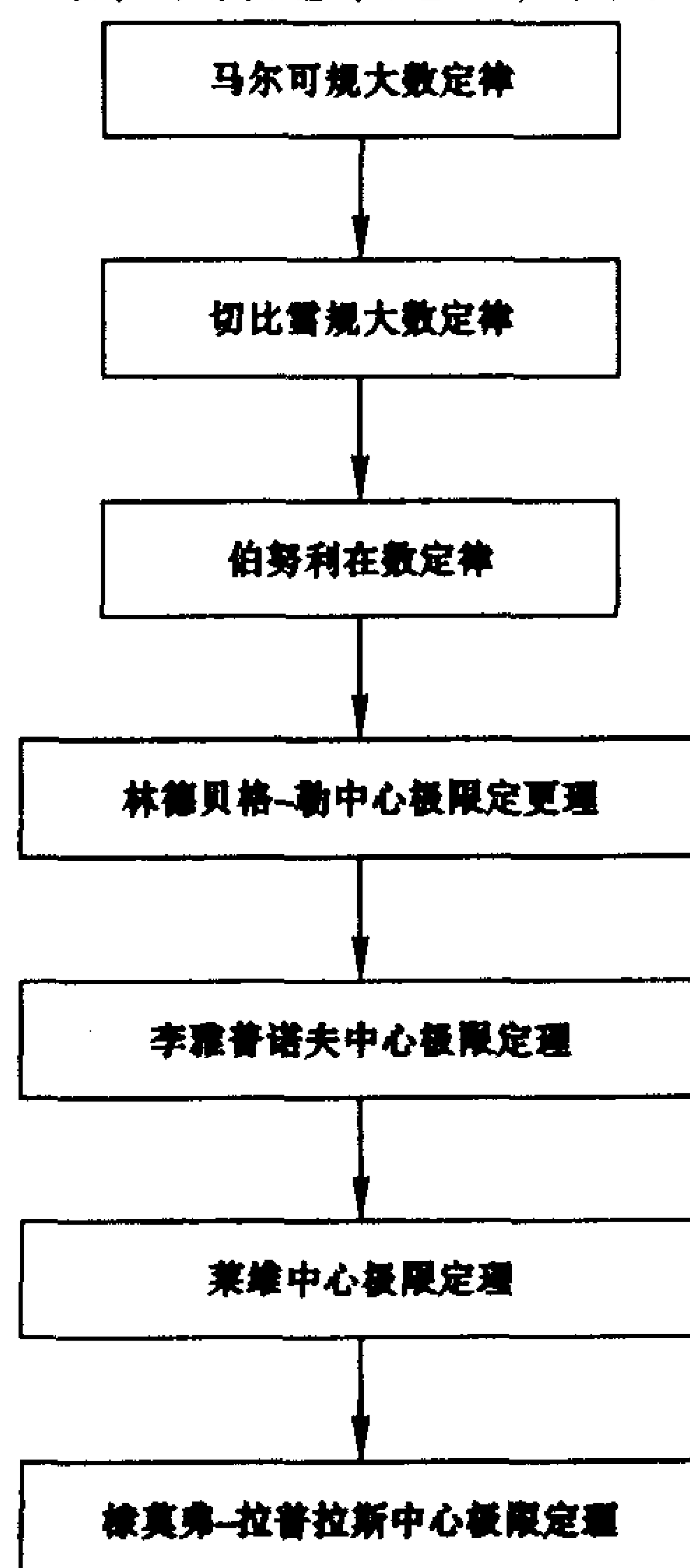


图 5-4

大数定律的成立条件略松，中心极限定理的条件较强，一般要求随机变量序列$\{X_i\}$相互独立，期望和方差均存在。中心极限定理中的李雅普诺夫条件为，存在$\delta>0$，使得

$$\lim_{n\to+\infty}\frac{1}{s_n^{2+\delta}}\sum_{i=1}^{n}E|X_i-\alpha_i|^{2+\delta}=0,$$

林德贝格-费勒中心极限定理的条件为

$$\lim_{n\to+\infty}\sum_{i=1}^{n}\sigma_i^2=+\infty,\lim_{n\to+\infty}\frac{\sigma_n}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}}=0。$$

显然，中心极限定理的条件较之大数定律要强一些。若是随机变量序列$\{X_i\}$独立同分布，则$\{X_i\}$同时满足大数定律与中心极限定理。关于这个事实，可以n个独立同分布的 0-1 分布合成的二项分布为例。设n_A为n次重复独立试验中事件A发生的次数，p为一次试验中A发生的概率，则有伯努利大数定律成立：

对任意给定的正数ε，有

$$\lim_{n\to+\infty}P(|\frac{n_A}{n}-p|<\varepsilon)=1。\qquad(5-11)$$

由于$\frac{n_A}{n}$这一比值为频率，所以伯努利大数定律的直观含义是：频率约等于概率，伯

努利大数定律构成了概率的统计定义的理论基础，概率论理论体系由此获得了首尾呼应。由此可见伯努利大数定律是相当重要的，正因如此，在此大数定律诞生两百周年的 1913 年，圣彼得堡科学院举行了隆重的纪念会。

1716 年由棣莫弗发现的中心极限定理、在 19 世纪初，拉普拉斯也发现了同样的事实，这就是概率论历史上最早的中心极限定理，即棣莫弗–拉普拉斯中心极限定理：

设随机变量序列 $\{X_n\}$ 服从二项分布 $B(n,p)$ ，则有

$$\lim_{n\to+\infty} P(\frac{X_n-np}{\sqrt{npq}}\le x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt \quad , \tag{5-12}$$

该极限定理建立了二项分布与正态分布的定量联系，打通了离散型与连续型随机变量的人为界限，为概率的近似计算，提供了一种方法。

在 $\{X_i\}$ 独立同分布、同期望、同方差的条件下，从中心极限定理可推演出大数定律。

事实上，

$$P(|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|\ge\varepsilon)=P(|\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}|\ge\frac{\sqrt{n}\varepsilon}{\sigma})$$

$$=P(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le-\frac{\sqrt{n}\varepsilon}{\sigma})+1-P(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le\frac{\sqrt{n}\varepsilon}{\sigma}) \quad ,$$

又因为

$$\lim_{n\to+\infty} P(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le x)=\Phi(x) \quad ,$$

所以

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|\ge\varepsilon)=\Phi(-\infty)+1-\Phi(+\infty)=0+1-1=0$$

第6章　概率论在中国的传播与发展

6.1 概率译名的历史演变

概率论（Theory of Probability）从19世纪30年代起就传入中国，但译名一直很乱。将英文“Probability”，俄文正式译为“概率”并广泛使用，是近几十年的事。1896年，（英）傅兰雅、华衡芳翻译概率论，书名《决疑数学》，用的是“决疑率”。1939年，段育华、周元瑞编写的《算学辞典》中将“Probability”译为“可遇率”、“或是率”。同年，（日本）长泽亀之助写的《代数学辞典》由薛德炯等翻译为中文，书中的概率被译作“或然率”。“或然率”这一译名在我国流行时间较长，直到解放初期仍在广泛使用。1948年，倪德基等编著的《数学辞典》，称概率为“适遇率”。1943年，赵缭编写的《数学辞典》中，概率又被称为“公算”。除此之外，还有“可能率”、“几率”、“盖然率”、“信率”等多种译名。日本数学界称概率为“确率”，由于中日两国文字相近，中国使用“确率”作为概率的译名的场合也比较多。概率的译名在所有数学术语译名中是最多的，但这些译名均抓住了概率是刻画随机事件发生可能性大小的这一实质，它们的出现为概率的译名的标准化作了准备。

1935年，国立编译馆编《数学名词》一书时，参考了众多的译名，将概率的译名缩小为“概率”和“几率”两个。解放后，中国科学院组织编写《数学名词》，在1956年的版本中，在1958年的《俄中数学名词》一书中都沿用“概率”和“几率”两名。这一工作，使译名逐步标准化。1964年，中国科学院编写了《数学名词补编》，确定使用“概率”为正式译名。1974年，中国科学院编写的《英汉数学名词》中，正式将“Probability”译为“概率”。

6.2 第一部在中国传播的概率论著作《决疑数学》研究

《决疑数学》的编译者华衡芳（1833~1902），字若汀，江苏常州金匮县（今无锡市）人，生于世宦之家，父亲华翼纶（?~1887）为举人，任永新知县。华衡芳7岁开始读书，14岁已读通程大位《算法统宗》飞归等题，对数学产生了很大的兴趣。他回忆十五六岁时“偶龄故书中检得坊本算法，心窃喜之，日夕展玩，不数月而尽通其义”。（华衡芳《学算笔谈》卷五）。其父见他这样喜爱数学，就从京城买回《周髀算经》、《九章算术》、《孙子》、《五曹》、《张邱建》、《夏候阳》、《缉古》、《海岛》诸算经以及《益古演段》、《测圆海镜》等供他攻读。华衡芳天资聪颖，刻苦自学，读书一年，无师自通。他为此感到自豪，说：“吾于算学，生平未尝受业于人;即与能算者相友善，亦未尝数数问难也”。以后他又读了秦九韶、梅文鼎、焦循、骆腾凤、李锐、罗士林、董佑诚等数学家的著作，以及《数理精蕴》、《几何原本》。

20岁前他遍览当时流行的古今中外数学书，奠定了雄厚的数学基础。其后华衡芳游学上海，结识了著名数学家李善兰。李氏比他年长22岁，向他介绍了正在同英国人威廉·亚历山大（Wylie Alexander，1815~1887）共同翻译的《代数学》和《代数积拾级》（1859），

告诉他："此为算学中上乘功夫，此书一出，非特中法几何尽废，即西法之古者，亦无所用之矣。"（《学算笔谈》卷五）在李善兰的引导下，华衡芳接触了代数学和微积分。这两部书出版后，李氏还送给他阅读。开始时他一点也不懂，"不得其用意之处"，则读几页，就已"不知其所语云何"，又请教李氏，回答说："此中微妙非可以言语形容，其法尽在书中，无所隐也，多观之则自解耳，岂是旦夕之工所能通晓者哉！"他按照李善兰的话去做，锲而不舍，反复钻研，从稍有头绪，终于豁然贯通。华衡芳 50 岁时回忆这一段自修心得，生动比喻说："譬如傍晚之星，初见一点，又见数十，数百点，以致灿然布满天空"。（华衡芳《学算笔谈》卷五）华衡芳一边努力吸收新学，一边进行研究和写作。他进取心强，年轻时见到李善兰的《火器真诀》后，觉得尚不能满意，二十五六岁就写出《抛物线说》。

1861 年，华氏 28 岁，学问已成，为曾国藩擢用，和金匮同乡好友徐寿（字雪村）一同到安庆军中，在金陵军械所一同绘制机械图，制造"黄鹄号"轮船，这是中国自造轮船的开端。是年冬，清廷命曾国藩保举人才，他推举了华衡芳、徐寿等 6 人，翌年他们被召到安庆府，在曾国藩幕中作幕宾。1863 年，曾国藩保奏华衡芳以县垂选用，受到洋务派重用。从此，他的一生便同洋务运动结下了不解之缘。

1865 年，曾国藩、李鸿章合奏创设江南制造局，华衡芳参加了这一新建工厂的计划和开创工作，"经始其事，擘划周详"。（《华衡芳家传》）1868 年江南制造总局内开设翻译馆，华氏与徐寿积极从事，出力较多。他们"志尚博通，欲明西学"，为介绍西方先进的科学技术，分门别类地进行系统译述。华氏与美国人玛高温等人共译《金石识别））、《地学浅释》等科学著作 5 种(另有两种未刊行)；与英国人傅兰雅((John Fryer, 1839~1929)共译《代数术》、《微积溯源》、《决疑数学》等著作 7 种（另有 3 种未刊行）。他从事译著十分辛苦，开始时稿本、改本、清本、草图等都出自他一手，夙兴夜寐，竟日操劳。华氏在上海居住三四十年，有为年译书，共刊行 12 种，170 余卷，内容上比李善兰的译著更加丰富，对清末传播西学产生了较大的影响。

华氏所译文字明白晓畅，内容丰富多彩，在李善兰之后介绍西方数学影响最大。华氏译著的特色是比较详尽地叙述了西方代数学史、三角学史、概率论史等。《决疑数字》"总引"以三千多字篇幅详述概率论史，提到的数学家约有 30 人，如迎但、耐普尔、费马、欧拉、拉格朗日、高斯等、提到了他们的数学成就和著作，为当时我国的学者打开了数理在翻译过程中，华氏继承了李善兰首创的数学符号和译名。这些几何学、解析几何、代数学和微积分的名词大都沿用至今，并传到日本。概率论名词华译流传下来的不多，有大数、指望溯望、排列、相关、母函数、循环级数等。华氏认为数学今胜于昔。他说："算法古疏今密，古拙今巧，亦天地自然之理"。"算学……后之视今犹今之视昔，安知此后更无再巧再密之术而视今之巧密者为疏拙耶?"这显然是对厚古薄今的质问。由于他坚信未来必胜于今天，曾预言"算器　滇盘、筹算等皆各有所长、亦各有所短，……吾意后世必有能创一器……而各事皆便者"。（华衡芳：《算草丛存》卷八《算斋琐语》）一百年后的今天已经普及的计算器，这恰好证明了他的这一预见的正确性。华氏的治学精神有两点特别值得称道。其一是："从来算学家皆喜炫其所长而匿其所短，余则反其所为，凡遇书中有未能之事与不能之事，必详言之，盖深望后之学者皆可以由此精进也"。其二是："古人之算学书每使人拌不易解，……病在于只将所造之境以示人，而其从人之途……皆秘匿而不道。· · ···余则力挽之习，于一切算法无不坦白以示人，……不求简奥，不避粗俗，唯使人易明而已。"（均《算斋琐语》）在提倡学术道德的今天，这两点仍值得我们效法。

综观华氏一生，26 岁开始从事数学研究和著述，30 多岁忙于办工厂、搞洋务，40 岁

以后从事译书、著作，是我国清末一位自学成才的数学家兼教育家。由于用心过度，积劳成疾，华衡芳先生67岁时患有脑病，70岁卒于家。夫人邹佩兰女士（1834~1873），为女诗人，11岁与华氏订婚，40岁时去世。生一子一女，子5岁夭殇，以弟子世芳为嗣。华氏虽官至四品，但非从政，一生不慕荣利，敝衣粗食，穷约终身，家无百金之蓄，实为我国近代数理统计和教育事业之先行者。

1.原本来源

在中国近代数学史研究中，关于传入中国的第一部概率论著作——《决疑数学》的原本问题，长期以来一直存在疑问。其中一种观点认为《决疑数学》的大部分内容来自《大英百科全书》第八版中伽罗威（Thomas Galloway，1796~1851）所写的《概率论》（A Treatise on Probability），而补充内容则来自《钱伯斯百科全书》中安德森的文章。这种说法得到了较广泛的认可。然而，经过详细的核对，发现《决疑数学》的唯一原本是伽罗威的文章，与安德森的文章无任何关系。

2.内容介绍

《决疑数学》是一部拉普拉斯概率论风格的著作，在总体结构上基本遵循了《分析概率论》的模式，包括一个3000字左右的“引论”，正文共有10卷，每卷有若干个问题（款），共160个问题。

《决疑数学》的“总论”。

“总论”部分非常简略，但较全面地叙述了在它出版的那个年代，即大略19世纪中叶，人们对于概率论这门学科的本质、概率论地应用价值及其发展历史的理解。关于伽罗威原文中的总引，著名的概率论史学家托德亨特在其概率论历史的巨著中曾经有过如下的描述：“这里我要提到在《珍藏本百科全书》（Cabinet Cyclopadia）中德摩根的关于概率论的文章的附录首页；在《大英百科全书》中成为托马斯·伽罗威的文章的总引的科学史；……这些会引起拉普拉斯工作研究者的兴趣。”的确，“总引”的整体结构完全按照拉普拉斯《分析概率论》总的引言的结构——关于概率的一般观点、概率论在各种具体事务总的应用总述、概率论发展历史的介绍展开的。

“总引”开宗明义，直接将以拉普拉斯为代表的概率观点明确地表示出来：“决疑数理为算学中最要之一门也。凡天下无一定之事，可先考其相关之故，而用算学推其分数之大小，以知其有否，此事之决疑数若何，或其事未必确定而心中疑信未定，则用决疑数可以自安其心。自古以来，格致家所考出之各种学问除数件事以足以自明之外（足以自明之事如几何之公论是也），其余各理未必果为真确，则有或多或少疑惑不决之处，或有千分之一、万分之一，不定之事皆可用决疑之理定其各事，约有若干分属于此，若干分属于彼。故于寻常习见之事用决疑数大有裨益。凡人心中所想各事必自觉其事之为是为否而揣度其事，所凭之故各分为两类，一为已知其成事之各源，有其各源定有其事，则从其各源以斟酌其事之是否可以告之;二为其成事之各源未知”。

卷一，“论决疑数之例”（第一~十款），计10款。

这一卷的主要内容来自《分析概率论》的第一章“理论的一般法则”。首先介绍了概率论中的基本概念，包括chance、probability、mathematical probability等，《决疑数学》将它们分别译为“不定”、“非的确”、“决疑率”等。本卷特别对概率这个基本概念的一些基本性质进行了详细的介绍。给出了拉普拉斯的概率古典定义

$$P(A)=\frac{k}{n},$$

其中n为基本时间的总数，k为事件A发生的次数。

关于概率的性质有以下几款：

第五款：$P(A)\le 1$，$P(A)+P(\overline{A})=1$。

第六款：对于两两互不相容的完全事件组$\{A_k\}$，有$P(\sum_k A_k)=\sum_k P(A_k)=1$。

第七款：对于独立事件的积（丛事），有$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)$。

第九款：对于互不相容的事件A_k，有$P(\sum_k A_k)=\sum_k P(A_k)$。

卷二，“论已试多次之事，其事为数件原事合成之事，其能有此事于不能有此事之各法已从决疑之理知之，而其数为常数者”。第十一~二十五款），计 15 款。

这一卷主要内容十讨论基于反复试验的事件或者由数个基本事件合成的事件的概率，即对于每一次试验，基本事件的概率十先验的÷不变的。给出了伯努利公式：

$$\frac{\text{一÷二三}\cdots\text{卯}}{\text{辛（）}}$$

即

$$b(k;\ n,p)=C_n^k p^k(1-p)^{n-k}=C_n^k p^k q^{n-k} \tag{6-1}$$

同时将二项概率推广到多项概率，即对于相互独立完全事件组$\{A_k\}$，$P(A_i)=p_i$，

$$\sum_{i=1}^{s} p_i=1,$$

那么在n此试验中，事件A_i出现k_i次的概率为

$$(p_1+p_2+\cdots+p_s)^n=\sum_{k_1+k_2+\cdots+k_s=n}\frac{n!}{k_1!k_2!\cdots k_s!}p_1^{k_1}p_2^{k_2}\cdots p_s^{k_s} \tag{6-2}$$

中含$p_i^{k_i}$的项。

卷三，“论连试多次之事或为任多原事合成之丛事，惟其能由此事之个法已能预知，而每试一次个法之数不同，求其事之决疑率。”（第二十六~三十款），计 5 款。

卷三内容事卷二内容的延续，不过卷三讨论的问题中，每一次试验基本事件的概率事先验的和变化的，例如无放回地摸球问题。着重讨论了超几何概率，即设共有N个元素，其中M个元素具有特征A，现从中非还原地随机取n个（$n\le N$），则恰好取到k个具有特征A的元素的概率为

$$h(k;\ n,N,M)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^k} \tag{6-3}$$

卷二和卷三的内容对应着《分析概率论》中第二章“由概率已知的简单事件复合而成的事件的概率”的内容。

卷四，“论决疑数之于景有关无关（谓关乎贫富之境与不关乎贫富之境也）二种，此两种之总谓之指望决疑数”（第三十款~四十一款），计 12 款。

此卷论述了拉普拉斯关于“数学期望”和“道德期望”的思想方法。其中的第三十四款是著名的“圣彼得堡悖论”。这个问题自从 1738 年伯努利在《圣彼得堡评论》发表的

文章提出来之后，由于这个问题的解决于概率论这一新兴的学科是否在实践中有效的问题有着密切的关系，因此吸引了众多的数学家的目光，如达朗贝尔、蒲丰、拉普拉斯、得摩根等。当时几乎所有关注过概率论的学者都曾经对这个有趣的问题进行过探讨。关于伽罗威在这一卷中对于“圣彼得堡悖论”的论述，卡尔·皮尔逊有过这样的评价：“迄今为止，关于它（圣彼得堡悖论）的最好的论述大概要属伽罗威在其概率论著作中给出的。”

卷五，“论从试验之事推算未来之事之决疑率”（第四十二款~五十二款），计 11 款。

此卷主要论述从已有的经验推导未来事件发生的概率。这部分主要是拉普拉斯的《分析概率论》第二卷第六章“关于原因的概率和从过去的经验所导致的未来事件的概率”的内容的一个简化。这里主要是以摸球为模型，进而推测将要发生事件的概率。第四十三款是最典型的描述：“设有一瓶内装有四球，其色为黑为白，不知其数，每取出一球，辨其色，仍还入瓶中，共试四次，得白球三次，黑球一次，求下一次再取一球其色为白为黑之决疑率”。第五十一款从有限的数目推广到无限的情形，导出了拉普拉斯的逆概率公式

$$\omega=\frac{P}{\sum P}=\frac{x^{m}(1-x)^{n}}{\int_{0}^{1}x^{m}(1-x)^{n}dx}.$$

也给出了全概率公式

$$P(B)=\sum_{i=1}^{n}P(B|A_i)P(A_i) \qquad (6\text{–}4).$$

卷六，“论籍人能治之决疑率以得利之各种题”（第五十三款~六十三款），计 11 款。

此卷讨论关于人寿保险的概率应用问题。根据一个人现在的年龄以及可能再生存年岁的概率计算生命保险金的问题，投保个人以此计算参加人寿保险每年应支付的保险费，人寿保险公司也以此保险业务的利润等。在《决疑数学》卷六中，人寿保险被翻译为“保命票”。拉普拉斯在其著作的第九章“基于将来事件发生的概率的收益问题”中专门讨论了这个问题。这一部分正好为快速发展的英国保险业提供了理论基础，因此，英国的数学家们开始对此投入了越来越多的关注。贝利、密利恩、得摩根和伽罗威等人对此做了大量的研究。他们还身体力行地参与此项实践活动，贝利在保险业挣了大笔的钱。

卷七，“论口证之决疑率并评理之十二人与状师定案之是非决疑率”（第六十四款~八十六款），计 23 款。

此卷是关于概率在法律诉讼方面的应用问题，论述古典概率论在证言÷陪审团和法庭判决中的应用。主要是《分析概率论》第二卷第十一章的内容的简化。关于证言的模型直接采取拉普拉斯的摸球模型，首先从最简单的包含两种颜色的球（对应着证言的真假）开始，然后是拉普拉斯对一个证人的证言的四种可能情况的分析，在推广到几个证人的情形。关于陪审团和法庭的判决的问题则主要采纳了泊松关于这方面的工作。

卷八，“论关乎极大之数各题之解法”（第八十七款~一百零九款），计 23 款。

此卷讨论了概率最大值问题与大数定理，证明了二项分布的最大值定理。是从离散组合的数学方法向连续的分析方法引进。就像拉普拉斯的《分析概率论》那样，从对于二项概率

$$b(k;\ n,p)=C_n^k p^k(1-p)^{n-k}=C_n^k p^k q^{n-k}$$

开始。

第九十款介绍了斯特林公式（《决疑数学》中称之为司林之例）

$$1\cdot2\cdot3\cdots x=x^{x}e^{-x}\sqrt{2\pi x}\text{。}$$

以上述公式为基础，从九十二款到九十六款都是对下面一般问题的推演：假设两个事件 A 和 B 各自发生的概率是 p 和 q，在试验 h 次中，A 发生 m 次，B 发生 n 次的概率是二项式展开式 $(p+q)^{m+n}$ 中第 $m+1$ 项。然后计算出 A 发生的次数处于 $hp\pm\tau\sqrt{2hpq}$ 的概率是

$$R=\frac{2}{\sqrt{\pi}}\int_{0}^{\tau}e^{-t^{2}}dt+P_{0}e^{-\tau^{2}}\text{，}\qquad(6\text{–}5)$$

其中 $P_0=\sqrt{h\div2\pi mn}$，$\tau=1/\sqrt{h\div2\pi mn}$。

这个式子与拉普拉斯的中心极限定理是等价的。这样就自然而然地推出了正态分布的函数

$$\int_{0}^{\tau}e^{-t^{2}}dt\text{。}$$

《决疑数学》第九十款说明了文章最后的正态分布表

$$\Theta=\frac{2}{\sqrt{\pi}}\int_{0}^{\tau}e^{-t^{2}}dt=1-\int_{\tau}^{\infty}e^{-t^{2}}dt$$

主要是根据克兰曾的《大气折射分析》中所制定的 $\int_{0}^{\tau}e^{-t^{2}}dt$ 改编的（用 $2/\sqrt{\pi}$ 乘即可）。

本卷此后的例子都是上述定理的应用。

卷九，“试论多次各次所得者不同所有决疑率最大之中位数并其有差之限之决疑率”（第一百十款~一百五十款），计 41 款。

此卷起始的第一百十款中说：“……所求者为试了多次而所试者非得或不得其事，但论物体大小，而其所大小在一定之限内有许多不同之同数或无穷不同之同数，各同数之决疑率或等或不等，而任意同数之决疑率每试一次不同。”显然，此部分内容对应着《分析概率论》第四章“关于大数次观察的平均结果的误差概率和最好平均结果的概率”。此卷对拉普拉斯这一部分误差理论的处理就像 19 世纪处理这部分内容的主流形式一样，作为进一步理解拉普拉斯理论的最初几个阶梯。

《决疑数学》此卷的前面 9 款（第一百一十三~一百二十二款）是选自泊松《关于犯罪和民事判决的概率的研究》的第四章，以及 1832 年的《天文年历》中的附录。泊松这部分的工作是十九世纪对拉普拉斯理论的最好的也是最重要的解释和发展。这卷后面的内容才正式介绍拉普拉斯在这方面的研究。

卷十，“论极小平方之法”（第一百五十一~一百六十款），计 10 款。

此卷是上卷的继续，主要是对拉普拉斯为代表的古典概率论中最有价值的成果之一——最小二乘法的介绍。本卷的大部分内容仍然是来自《分析概率论》第四章，不过《决疑数学》卷十中的最后 4 款，即第一百五十七款至的一百六十款，对于最小二乘法的处理方法则采纳了高斯在《使综合观察误差尽可能小的理论》中的方法。

3.历史评价

傅兰雅与华衡芳合译的《代数难题解法》中已有概率论的知识，它是西方概率论知识第一次传入中国，但是其内容并不系统全面。《决疑数学》包括了概率论学科中的最基本、最具代表性的内容，因此是一部开始在中国传播的系统而全面的概率论著作。《决疑数学》的出版具有重要的意义，它不仅是概率论这一新的近代思想传入中国之始，而且就像拉普拉斯在 19 世纪对整个欧洲社会合科学所产生的影响和冲击一样，它也为古老的中国社会带来了一丝清新的空气。

但是，《决疑数学》在中国问世之时，也是拉普拉斯概率理论在欧洲数学界引起广泛争论的时期。

作为第一部中文概率论著作，华衡芳与傅兰雅要做的第一件事是必须找到或者发明适当的中文词汇与概率论中的术语和词汇相对应。他们将概率论中的一些重要术语，如 probability（概率）、event（事件）、simple event（基本事件）、compound event（复合事件）、independent cases（事件的相互独立性）、trial（试验）、expectation（期望）、mathematical expectation（数学期望）、moral expectation（道德期望）、error（误差）、weight（权）、average error, or mean error（平均误差）、the method of least squares（最小二乘法）、generating function（生成函数、母函数）等分别翻译成决疑率、事、原事、从事、彼此决不相关、试指望、不关景决疑数、关景决疑数、差重数、中差、极小平方之法、母函数等。现在看来，《决疑数学》中对于概率论中术语的翻译大多数没有沿用下来。

其次，华衡芳等仍然沿用了李善兰开创的西方数学的翻译风格——用汉字代替西方数学符号和数字。如阿拉伯数字 1、2、3、4、5 译为中国数字一、二、三、四、五，小写拉丁字母 a,b,c 译为甲、乙、丙，未知数 x,y,z 译为天、地、人。在这样一套符号体系下，概率论中的许多命题和公式表述起来极为复杂和难懂。

尽管从整体上来讲，译者对此书的内容的翻译细致的，内容的把握也比较准确，然而其使用的仍然是中国传统数字和记号，这对于概率论这样一门需要极其复杂的高等数学知识的数学学科而言，显然大大地不合时宜了。近代数学的本质特点之一就是符号化和抽象化，《决疑数学》的翻译风格除了平添了内容表述上的难度外，也增加了阅读它的人对于概率论本质、内容理解的难度。

因此，对于概率论在中国的传播和普及，就某种程度而言，《决疑数学》并没达到其应有的作用。

6.3 许宝騄等中国数学家的工作

许宝騄（1910~1970）是中国数学家。1910 年 9 月 1 日生于北京，1970 年 12 月 18 日卒于北京。

许宝騄祖籍浙江杭州，出生于名门世家，少年时代的许宝騄受益于表姐夫徐传元（毕业于美国麻省理工学院）的指导。1928 年毕业于北京汇文中学，毕业后先考入燕京大学理学院，后来了解到清华大学数学系最好，自己又对数学兴趣最浓，于是 1929 年转入清华大学攻读数学，1933 年获理学学士学位。毕业后经考试被录取赴英留学，但由于体重太轻不合格未能出国，然后到北京大学数学系当助教。1936 年他再次考取了赴英留学，在伦敦

大学当研究生，同时在剑桥大学学习，1938 年获哲学博士学位。1940 年又获科学博士学位，同年回国，任北京大学教授，执教于昆明西南联合大学。1945 年再次出国，应邀先后在美国伯克利加州大学，哥伦比亚大学和北卡罗来纳大学任访问教授，1947 年回到北京大学任教。1948 年当选为中央研究院院士，1955 年当选为中国科学院学部委员。

许宝騄积极倡导学科振兴，热心培养人才，仅在北京大学就培养了 8 届概率统计专门化学生，亲自指导了 5 届学生的讨论班和毕业论文。特别是他晚年在身体很不好的情况下，在北京大学同时领导了数理统计、马尔可夫过程、平稳过程三个讨论班，希望把一批年轻人带到科研的前沿。他讲课深入浅出，一个复杂的问题经他分析后变得明白自然。近 20 多年活跃在国内外的不少著名数理统计和概率论领域的学者、教授都是他培养的学生。

许宝騄学习非常勤奋、刻苦。例如，在昆明西南联大时，生活清苦。资料贫乏，那时找一本书都困难，他曾手抄过梯其马舍的整本《函数论》，他念过的书，往往都写了不少批注，有的书都被他翻得成零页了。他在学术研究方面，知难而进，积极参与重大问题的探索。他总是寻求简明、初等的方法，他认为初等方法比艰深的方法更有意义。他追求一个问题的彻底解决，追求一般性。他一生未婚，长期带病工作。他晚年已瘫痪，卧床不起，让人借来“文革”期间出版的全部《数理统计纪事》，两个月内顽强地阅读了几年的杂志，了解到当时的情况，写下了他最后一篇论文。1970 年 12 月他逝世时，床边小茶几上仍放着钢笔和未完成的手稿。

许宝騄的上述精神和品格深深地感动着他同行和学生。例如他的学生和同事著名数学家安德逊、钟开莱、莱曼在一篇他们共同写的文章中说：“许（宝騄）坚持深入浅出，毫不回避困难。特别是沉着、明确而默默地献身于学术的最高目标和最高水准，这些精神吸引了我们。”

1981 年，著名的施普林格出版社，刊印了由杰出数学家钟开莱主编的《许宝騄全集》。1984 年，由徐利治、郑清水和他的学生钟开莱发起设立了以他的名字命名的统计数学奖—“许宝騄统计数学奖”，奖励 35 岁以下研究数理统计与概率统计的青年工作者，这是中国最高的数学奖项之一。

1880 年，英国学者傅兰雅和中国数学家华蘅芳编译的《决疑数学》是传入我国的第一部概率论著作。由于种种因素，该书对我国的概率论发展没有产生多大影响。辛亥革命后，微积分、近世代数、近世几何学等相继进入中国的高等教育领域，而概率论尚未进入。1915 年 1 月创刊的中国第一份现代科学杂志《科学》曾刊出一篇文章《最小二乘式》，此为中国第一篇概率论文章。后胡明复（1891~1927）曾撰写《几率论》、《误差论》等一系列论文探讨概论统计的哲学问题。由于受中国传统数学思想的影响，加之近代数学基础薄弱，随机数学在中国发展甚是缓慢。直到 20 世纪 30 年代，中国数学家褚一飞、刘炳震、许宝騄、钟开莱等才陆续发表概率论与数理统计的研究论文，拉开了中国对概率论与数理统计研究的序幕。

许宝騄是 20 世纪中最富有创造性的统计学家之一，是中国最早在概率论与数理统计研究方向达到世界先进水平的杰出数学家。他加强了强大数定律；研究了中心极限定理中误差大小的精确估计；发展了矩阵变换技巧；得到了高斯-马尔可夫（Gauss-Markov）模型中方差的最优估计；揭示了线性假设似然比检验的第一个优良性质等。其研究成果已经成为当代概率论与数理统计理论的重要组成部分，至今“许方法”仍被认为是解决检验问题的最实用方法。

少年时代的许宝騄受益于表姐夫徐传元(毕业于美国麻省理工学院)的指导。1928 年，许宝騄考入燕京大学化学系，但对数学的浓厚兴趣，促使他改攻数学，并于 1930 年考入清华大学数学系。期间，深受熊庆来（1893~1969 年）、孙光远（1900~1979 年）和杨武之（1896~1973 年）的教诲。1933 年，以优异成绩获得理学士学位。1936 年，通过赴英庚子赔款公费留学考试，进入伦敦大学学院（University College）的高尔顿（ F.Galdon，1822~1911）实验室和统计系学习数理统计学。1938 年获得哲学博士学位，两年后又获得理学博士学位。

1940 年，许宝騄回到抗日烽火中的祖国，受聘为北京大学教授，在西南联合大学任教。1945 年，应加州伯克利大学和哥伦比亚大学的联合邀请而前往美国。1947 年 10 月，谢绝众多朋友的挽留，毅然回到中国，此后一直在北京大学任教。

许宝騄是中央研究院第一届当选的 5 名数学所院士之一。1955 年当选为中国科学院学部委员。1979 年美国《数理统计学年鉴》高度评价了他对概率论与数理统计学科所做出的卓越贡献。1981 年和 1983 年，科学出版社和德国施普林格（Springer-Verlag）出版社分别出版了《许宝騄文集》和《许宝騄选集》。在美国斯坦福大学统计系走廊里至今悬挂着许宝騄的画像。1984 年，为了纪念许宝騄及推进我国统计学的发展，数学家钟开莱、郑清水、徐利治发起设立“许宝騄统计数学奖”，奖励 35 岁以下研究数理统计与理论统计的青年工作者。这是我国最高的数学奖项之一。

许宝騄痛感中国数学之落后，怀着满腔的报国热情，决心把自己的事业立足于祖国。由于概率论与数理统计在中国几乎是空白的学科领域，于是，许宝騄以惊人毅力和无私奉献精神为其奠定了基础，并为之振兴付出了毕生精力。

在实际工作及理论问题中，概率接近于 1 或 0 的随机事件具有重要意义。概率论的一个基本问题就是探索概率接近于 1 的规律，特别是大量独立或弱相依因素累积结果所发生的规律。大数定律就是研究这种规律的命题之一。许宝騄对大数定律进行了深入探讨。

强大数定律和弱大数定律取决于收敛的类型。第一个弱大数定律由雅各布·伯努利提出，刻画了大量经验观测中呈现的稳定性。后泊松又提出了一个条件更宽的陈述，即泊松大数定律。切比雪夫第一次严格地证明了伯努利大数定律，并把结果推广到泊松大数定律。1866 年，切比雪夫给出著名的切比雪夫不等式，并由此导出切比雪夫大数定律。

第一个强大数定律由法国数学家博雷尔在 1909 年对伯努利试验场合建立。他证得若试验次数无限增加时，频率将趋于概率。博雷尔的工作激起了数学家沿这一崭新方向的一系列探索， 其中尤以柯尔莫戈洛夫的研究最为卓著。他在 1926 年推导了弱大数定律成立的充分必要条件，后又对博雷尔提出的强大数定律给出了一般结果。

许宝騄进一步加强了强大数定律的结论，其结果为：

设 X_1，$X_2,\cdots X_n,\cdots$ 是独立同分布均值为零、方差有限的随机变量序列，任给 $\varepsilon>0$，有

$$\sum_{n=1}^{+\infty}P(\frac{1}{n}|X_1+X_2+\cdots+X_n|>\varepsilon)<\infty \quad (6\text{-}6)$$

证明是经过一个卷积的富立叶逆转，把问题转化为含有特征函数某个积分的分片估计，这需要具有相当深厚的数学功底和敏锐的数学眼光才能完成。由于推证较复杂，尽管已经得出关于矩的充要条件，但在刊出时删去了必要性的证明。

概率论中的极限定理研究的是随机变量序列的某种收敛性，对随机变量收敛性的不同

定义将导致不同的极限定理。许宝騄在“依分布收敛”、“依概率收敛”、“r 阶收敛”和“依概率 1 收敛”的基础上，创造性地提出“完全收敛性”概念，开辟了概率论极限理论研究的新局面。直到今天，对完全收敛性的讨论仍是一个有意义的课题，这就足以表明该文的开创性价值。正如许宝騄所说：“一篇论文不能因为获得发表就有了价值。其真正价值要看发表后被引用的状况来评价。”

许宝騄对中心极限定理也进行了较为深入的研究。

“中心极限定理”这个术语是由波利亚（G.Polya，1887~1985）1920 年引入的。该定理断言在适当条件下，大量独立随机变量和的概率分布近似于正态分布。在长达两个世纪的时间内极限定理成了概率论的中心课题。

1733 年，棣莫弗由二项分布的渐进分布推导出正态分布。较一般的极限定理由拉普拉斯给出，但其证明不完善。

误差分析是概率论的生长点之一。如果把随机变量总和中的每项看做是小的“基本误差”，那么中心极限定理就为观察误差中正态分布的发生给出一个解释。19 世纪初高斯在研究测量误差时引进了正态分布，并发展了具有广泛应用的最小二乘法。

在许多数学家为给出中心极限定理严格证明所做的努力均告失败后，切比雪夫使用矩方法的尝试相当令人鼓舞。马尔可夫于 1887 年第一个用矩方法给出了中心极限定理的严格证明。切比雪夫的另一个弟子李雅普诺夫则从一个全新角度去考察中心极限定理，引入特征函数这一有力工具，避免了矩方法所要求的高阶矩存在的苛刻条件，在 1901 年给出了定理的完善证明，其证明方法与现在素数理论中的方法相类似。特征函数实现了数学方法的革命，为极限定理的进一步精确化提供了条件。

一个从理论和应用上都应当关心的问题是，仅知道某个概率分布渐近正态分布是不够的，还必须知道换成正态分布后误差有多大。李雅普诺夫给出这个误差的一个上限。瑞典数学家克拉美（H.Cramér，1893~1985）发现李雅普诺夫所给余数的估计在风险问题中是远远不够的，并于 1928 年改进了结果。1941 年，贝莱（A.C.Berry）再次改进了李雅普诺夫的结果。

许宝騄有一本翻破了的克拉美概率著作，书上几乎写满了批注。他认为该书包含了所有概率论的基础。1945 年，许宝騄改进了克拉美定理和贝莱定理，并给出克拉美定理的一个初等证明。他以特征函数为工具，通过 12 个引理，给出了上述定理的证明。但影响更深远的结果是他将相应的样本均值代之以样本方差。

许宝騄说：“关于均值的渐近分布，已知结果如此之多。考尼斯（Cornish）和费希尔通过半不变量获得了逐步近似于任何随机变量分布的各项。若把考尼斯和费希尔的形式结果转化为一条渐近展开的数学定理，它能给出剩余项大小的阶。在本文中，样本方差就做到了这一步。”

这里许宝騄第一个讨论了样本方差的渐近展开，给出余项阶的估计。他直接引进了一个新维数，用特征函数来近似随机向量的分布，其难点是用特征函数来近似两个高度相关的随机变量的分布。他对特征函数的应用已经达到炉火纯青的境界，在不少论文中对这一技巧信手拈来，应用自如。

许宝騄所采用的方法具有普遍意义，还可以用于解决样本高阶中心矩、样本相关系数及样本统计量的类似问题。他的这一工作在 20 世纪 70 年代以后引起了进一步的研究条件。

1947年5月，他得到每行独立的无限小随机变量三角阵列的行和，依分布收敛于此后，许宝騄开始研究费勒（W.Feller，1906~1970）对中心极限定理一般形式的充要一给定的无穷可分律的充要条件。当时一些著名的概率专家，如柯尔莫戈洛夫、辛钦、格涅坚科（B.V.Gnedenko，1912~1995）、莱维（ P.Lévy，1886~1971）和费勒等，都在寻找这一答案，所以许宝騄在给钟开莱的信中说，担心正在进行的工作会和别人相重复。

许宝騄的条件与格涅坚科的不同，后者的“两个尾巴”是并在一起的，而许宝騄则利用核$(\sin t/t)^3$直接证明。但得知格涅坚科的研究成果已经发表时，许宝騄立即承认了其优先权。因此，在格涅坚科和柯尔莫戈罗夫合著的相关专著英译本再版时，添加了许宝騄的这一论文作为附录。

20世纪50年代中期，许宝騄对马尔可夫过程产生了兴趣，他用分析的方法讨论了关于转移概率函数的可微性。这一工作暗示了分析结构和概率结构的内在联系，为进一步研究奠定了基础。

王梓坤（1929~）出生在湖南省零陵县的一个家境极其贫困的家庭，从小缺衣少穿。但他从中学起就兴趣广泛，很注意方法问题。1952年毕业于武汉大学数学系。1955年在南开大学任教期间考取留苏研究生，在莫斯科大学数学力学系攻读概率论专业，其导师为现代概率论的奠基人柯尔莫戈洛夫和很年轻且富有才华的杜布罗辛。导师对他的耐心指导，至今令王梓坤心存感激。1958年，王梓坤的论文《生灭过程的分类》在莫斯科大学学术答辩会获得一致通过，被授予副博士学位。

王梓坤学成回国后教学和科研的主要方向在概率论，早在60年代他就是我国概率论的领袖人物，为我国概率论的教学和科研做出了巨大的贡献。

王梓坤先生是中国概率论研究的先驱和学科带头人，主要工作如下：

（1）首创极限过渡的概率方法，彻底解决了生灭过程的构造问题。

（2）1961年，首创用差分方法研究生灭过程泛函的分布以及停时与首达时的分布，得到了深入的结果。

（3）关于马尔可夫过程一般性质的研究。

（4）1980年以后，研究马尔可夫过程与位势论的联系，发表了论文《布朗运动与位势论（上）（下）》、《布朗运动的末遇分布与极大游程》、《对称稳定过程与布朗运动的随机波》和专著《布朗运动与位势》。1983年以后研究多指标马尔可夫过程。

（5）除马尔可夫过程的研究外，王梓坤还开创了随机泛函分析在中国的研究。在他的带动下，中国目前在这方面的工作已经很多。

（6）研究多指标马尔可夫过程在中国最早，在国际上最早引进多指标奥恩斯坦——乌伦贝克过程的定义，并取得了较系统的成果。

（7）发表《生灭过程与马尔可夫链》等专著9部。其中《概率论基础及其应用》（1976）、《随机过程论》（1965）和《生灭过程与马尔可夫链》（1980）三部著作从概率论的学科基础到研究前沿构成完整体系，对中国概率论的教学和科研起了非常重要的作用，被一些名牌大学用作研究生和本科生的教材。

（8）在概率论应用方面研究的最大成果是，对地震的统计预报以及与军队合作完成了在计算机上模拟随机过程的研究。

陈希孺（1934~2005），1934年2月11日生于湖南省望城县，1956年毕业于武汉大学

数学系，后就职于中国科学院数学研究所。1961年到中国科技大学从事数理统计教学与研究工作至今。陈希孺1980年任教授，1981年任博士生导师，1997年当选为中科院数理学部院士，现担任中国数理统计学会理事长及中国统计学会副会长等职。是中国当代概率论与数理统计专家，中国科学院研究生院教授，曾任中国数学会概率统计学会理事长，1997年当选中国科学院院士。陈希孺在建立中国现场统计研究会和中国概率学会中起了重要的作用，并担任现场统计研究会的理事长、中国统计学会副会长，国家技术监督局全国统计方法应用标准化技术委员会的委员兼一分委主任，主持和参与制定了多项有关统计应用的国家标准。

陈希孺先生研究领域主要为：线性模型、U统计量、参数估计与非参数密度、回归估计和判据等数理统计学若干分支。

主要成果：

（1）对线性统计模型作深入系统的研究，圆满地解决了一般损失函数下M估计的强、弱相合问题。

（2）在非参数统计量，特别是极重要的U统计量的研究中获得U统计量分布的非一致收敛速度，具有国际领先水平。

（3）在非参数回归、密度估计与判别中做出了一系列优秀成果，包括定出了错判概率的指数限，“data based”型估计的收敛条件，以及对几个常用的密度估计和回归估计类定出了最佳收敛速度等。

（4）在参数估计这个基本分支中，解决了国际统计学界当时致力的一些问题，包括定出了重要的正态分布两参数在一般损失下的序贯Minimax估计，否定了关于某种区间估计存在条件的一个公开猜测，并提出了正确解释。

陈希孺有专著和教科书11部，在国内外重要刊物上发表论文100多篇。代表作为教科书《概率论与数理统计》（1992）、专著《机会的数学》（2000）。曾获得中国科学院中大科技成果一等奖，中国科学院自然科学一、二等奖，国家自然科学三等奖，中国科学院教学成果一等奖和贵州省科技进步二等奖。

6.4 历史的反思

关于概率论的历史，几乎所有教本都是一带而过，很少有详尽的论述，这是应该弥补的不足。任何一门学科的草创、形成以及发展都有其连续性，割断历史、忽略历史都会给教学和科研造成不应有的损失。

概率论的历史可分作两部分来阐述。

1.国外史

国外史可借助于表6-1作出详尽的说明。表中的人物是概率论发展史上具有里程碑意义的人物。这样一来，表格被分割成为国名、人名、生卒年份、主要贡献四个区域，这四个区域分别可以说明四个重要的历史问题。

表6-1

国名	人名	生卒年份	主要贡献（或著作）
法国	帕斯卡	1623~1662	点数问题
	费马	1601~1665	

续表 6-1

国名	人名	生卒年份	主要贡献（或著作）
荷兰	惠更斯	1629~1695	《论赌博中的计算》
瑞士	雅各布·伯努利	1654~1705	《猜度术》
法国	棣莫弗	1667~1754	《机遇论》
法国	拉普拉斯	1749~1827	《分析概率论》
德国	高斯	1777~1855	正态分布
法国	泊松	1781~1840	泊松分布
俄国	切比雪夫	1821~1894	《论均值》
苏联	柯尔莫戈洛夫	1903~1987	《概率论基础》

（1）概率论的起源。

列于国名这个区域的都是欧洲的大陆国家，其中法国出现的频率最高，因此可以得知，概率论起源、形成于欧洲的大陆国家，发展的中心最早在法国，随着时间的推移中心在东移，先从法国移到德国，又从德国移往俄国。这种推断是正确的，二次大战后，概率论的中心已在美国登陆。

（2）概率论的命运。

在人名这一区域内，所有为微积分的创立和奠基立下汗马功劳的数学大师们，如牛顿、莱布尼兹、柯西、外尔斯特拉斯等，一个都不在内，这是否暗示着概率论的命运与微积分截然不同呢?

众所周知，微积分创立后，先后由无穷小和悖论问题引起了两次数学危机，几代数学家、哲学家甚至革命家（如马克思）在几百年的时间内都卷入其中争执不休。而概率论呢?则完全是在不闻不问、听之任之的气候中慢慢成长起来的。那么，是什么样的一种机制在起作用呢?这便是下面接着讨论的问题。

（3）概率论的发展机制。

再看时间这一栏目，帕斯卡在世不到 40 年时间，在他 20 岁的时候，发明了世界上第一台机械计算机，在他 30 岁时，首先用概率论的方法解决了赌博中的若干问题。大家知道，在帕斯卡 20 岁到 30 岁这段时间内，英国确立了资本主义制度（1649 年 5 月），中国的末代王朝在北京定都。看来，概率论是随着资本主义制度的确立慢慢成长起来的，那么为什么概率论没在清代出现，而偏偏在建立资本主义制度的地方出现呢？看来在封建社会制度中也不具备有利于概率论发生、成长的内在机制。那么，这种机制到底是什么呢?这个问题在讨论完第四区域后便可明缭。

（4）概率论的逻辑起点。

若把第三、四两个区域对照起来看。将出现一个巧合的现象：人物的生死次序与概率论的次序刚好相反！也就是说，如果要写概率论的发展史的话，需从帕斯卡到柯尔莫哥洛夫，但是要学习概率论的话，最好还是从柯氏讲到帕斯卡。概率论的逻辑次序按人名可排成图 6-1 形式。

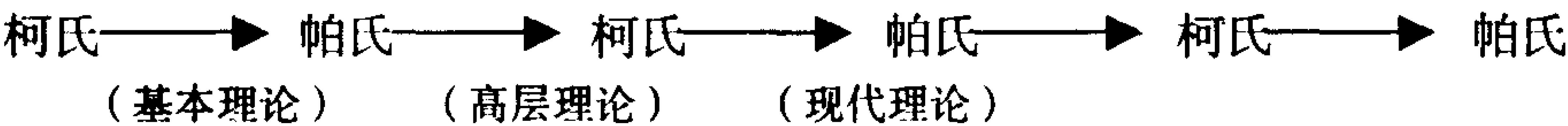


图 6-1

这里有一个明显的循环（人名）递进（理论）现象，循环递进的起点是柯尔莫哥洛夫的公理化体系，因此公理化定义是概率论的逻辑起点，大学理工科的概率论教学不妨从此开始。根据（4）的分析，可以回答（2）、（3）提及的概率论的发展机制问题，这种封建制度内部不具备的、而资本主义制度内部具备的机制，就是“循环递进机制”，这种机制的催化剂是实验科学。从这一点出发，便可得知概率论没有在清初出现的原因，乃是历代封建王朝（特别是清代）轻视实验科学的开发和研究，这几乎是封建王朝的不治之症。因而清朝虽然一直存在着赌博现象，但是从来也没有人把它上升到科学的高度，所以概率论根本不可能在中国首先出现。退一步讲，即使首先在中国出现，也会因缺乏实验科学这一催化剂而无法建立起概率论的循环递进机制，以致最终发展不起来。在这方面，平面几何就是个教训。

2.国内史

1880年，曾国藩的弟子、洋务运动的大将华衡芳（1833~1902）在上海翻译了第一本概率论书籍，名曰《决疑数学》，由于华衡芳经常到天津讲学，因此不久就传到了天津;1896年，《决疑数学》出版后，很快传到京师大学堂，即今天北大和北师大的前身。抗战胜利后（1947年10月），许宝缘博士回国在北大任教，概率论的重心基本上移到了北京。解放后，尽管王梓坤先生在柯尔莫戈洛夫的门下学成归来，在南开大学任教，但并没有改变概率论重心偏移的走向。1970年，许宝騄教授去世后，概率论的重心随侯振挺到达了长沙;1980年后，我国的概率论的研究及应用又在上海形成了以华东师大、复旦大学为核心的新的中心。

这样，在大约100年的时间里，概率论在我国也出现了“循环递进”现象，有没有催化剂呢?严格地讲是有的！这便是1937年毛泽东在延安写的《实践论》，毛泽东在《实践论》中提出了认识——实践——再认识，再实践这种具有循环递进机制的理论模式，这是马克思主义发展史上的第一次。但是由于人所共知的原因，并没有很快形成科学实验的学说，因而也就很难指导实验科学，因此对概率论的发展帮助不大。

对比概率论发展的两部历史不难发现，概率论实践论的核心，就在于建立起“循环递进”的机制，对这种机制应该做广义的理解。在概率论的教学与实践中，若能建立起一种“循环递进”的机制，则可使我们获得一盏神灯。人们在教研中，最容易出现、最容易满足、最难以突破的就是胶着的“循环”状态，若想冲破循环的死锁，则必须建立一种“递进”的机制，至于如何建立这种机制，则往往因人而异。

正是这种差异性，才造成了概率论不同学派纷争的多姿多彩性。概率论历史、现状以及未来的迷人魅力和广阔前景，亦概源于此。

西方概率论在中国的早期传播，主要是通过中国学者的翻译和编译的形式进行的。尽管西方的概率论教科书被大量翻译或编译过来，但作为一种数学文化或价值观，古典概率论中所包含的机会均等思想似乎并没被国人普遍认识和接受。上个世纪，虽然中国的数学事业经历了曲折的道路，但在广大数学工作者的共同努力下，仍获得了巨大的成绩，随着中国数学发展水平与国际地位的不断提高，中国数学家在概率论领域的研究也必将取得更大的成就。

第7章　概率论发展中若干问题的思考

困扰科学史界的一个问题是随即现象的研究很早就有，为什么直到17世纪才正式成为一门学科？二是在早期中国为什么没有出现概率理论？

7.1 概率论起源的哲学思考

长期以来，概率论一直被认为是从赌博游戏中产生的。但是考古学发现告诉我们，赌博游戏早在文明初期就已经存在了，迄今已有几千年的历史，而概率论却直到17世纪末才诞生，至多不过三百多年历史，这说明赌博不是概率论产生的决定性因素，在从赌博出现到概率论产生之间的这段“空白”期，必定还有一些十分关键的因素正在孕育之中，概率论的形成是这多种因素结合的结果。

对概率论而言，两个最主要的概念就是独立性和随机性。概率论是从研究古典概型开始的，它所涉及的研究对象是大量的独立随机过程。通过对这些过程中出现的问题的解决，概率理论体系才逐渐地建立起来。因此要考察概率论的产生条件，首先应对独立随机过程的产生有充分的了解。

事实上，这种过程的雏形早在原始社会就已经存在了，那时的占卜师们使用动物的趾骨作为占卜工具，将一个或多个趾骨投掷出去，趾骨落地后的不同形状指示神对人事的不同意见。由于投掷趾骨这个过程所产生的结果具有不可预测性，而每次投掷的结果也互不影响，这与我们今天投掷骰子的基本原理相当，因此趾骨可以被看做是骰子的雏形。但是由于趾骨形状的规则性较差，各种结果出现的几率不完全相同（即不具有等可能性），所以趾骨产生的随机过程还不是我们今天意义上的独立随机过程。加之趾骨作为一种占卜工具，其本身具有神圣的地位，普通人不可能轻易使用，这也在一定程度上阻碍了人们对随机过程的认识。

随着社会的进步和文明的发展，骰子变得越来越普及，不仅数量增多，规则性也日益精良，此时它不再是一件神圣的器具而逐渐成为普通大众的日常用具。从原理上看，只有一枚骰子是质地均匀的，它就可以产生一系列独立的随机过程。哲学过程具有良好的性质（独立性、随机性、等可能性），是进行概率研究的理想对象。如果经常接受这些随机过程，就有可能从中发现某些规律性。实际上，通过对骰子的研究我们确实发现了有些有趣的现象。在考古出土的骰子当中，有一些被证明是用于赌博的工具，它们的形状规则而质地却不均匀，也就是说，如果骰子的某一面较重，则其对面朝上的几率就会增大，这种骰子明显是为了赌博时用于作弊。而从另一个角度来看，如果古代人知道质地不均匀的骰子产生各个结果的可能性不同，那么他们必定清楚一个均匀的骰子产生任何一个结果的几率时相等的。也就是说，经常从事赌博的人必然可以通过大量的游戏过程，意识到掷骰子所得到的结果具有某种规律性，并且这种规律性还可以通过改变骰子的质地而得到相应的改变。虽然古代人的这些意识还只停留在经验的总结的水平上，却不得不承认这是一种最原始的概率思想。

赌博游戏存在的时间之长、范围之广、形式之多令人惊讶。但有如此之多的人沉迷于

这种游戏，也在客观上积累了大量的可供学者进行研究的随机过程。更为重要的是，在赌博的过程中，或许是受到经济利益的驱使，已经开始有人试图解开骰子的奥秘。意大利数学家卡尔达诺就是其中的一位。他本人是个大赌徒，嗜赌如命，但他具有极高的数学天分。在赌博的过程中，卡尔达诺充分发挥了他的数学才能，研究可以长胜不输的方法。据说，他曾参加过这样一种赌法：把两颗骰子掷出去，以两颗骰子朝上的点数之和作为赌的内容。那么，赌注下在多少点最有利?

$$\begin{matrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 8 & 9 & 10 & 11 \\ 7 & 8 & 9 & 10 & 11 & 12 \end{matrix}$$

两颗骰子朝上的面共有 36 种可能，点数之和分别为 2~12 共 11 种可能，从上图可知，7 位于此六阶矩阵的对角线上，它出现的概率为 6/36=1/6，大于其它点数出现的概率，因此，卡尔达诺预言押 7 点最好。这种思想今天看来很简单，但在当时却是杰出的。他还以自己丰富的实践经验为基础，写成了全面探讨赌博的《机遇博弈》一书，书中记载了他研究赌博的全部成果，并且明确指出骰子应为“诚实的”（honest），即六个面出现的机会相等，以便在此基础上研究掷多颗骰子的等可能结果数。

这些实例充分说明，赌博曾对概率论的产生起过积极的作用。这可能就是人们在谈到概率论时总是把它与赌博联系在一起的缘故吧。但是，我们应该认识到，赌博的价值并不在于其作为一种游戏的娱乐价值，而在于这种机遇游戏的过程实际上就是良好的独立随机过程。只有出现了独立随机过程，概率论才有了最初的研究对象。而概率论也的确是在解决机遇游戏中出现的各种问题的基础上建立起自己的理论体系的。因此，在概率论的孕育期，可以作为一种模型进行研究的机遇游戏过程即独立随机过程的出现是概率论得以产生的一个重要的全体条件。

前面曾经提到，独立随机过程的出现并不是概率论诞生的决定性因素。仅有概率思想而不能将概率结果表达出来，也不能形成完整的理论。概率论是一门以计算见长的首先分支，计算过程需要运用大量的加法和乘法原理（组合数原理）进行纯数学运算。对于现代人来说，概率计算并不是一件难事。但是对于 16 世纪以前的人来说，计算却是十分困难的，原因就在于古代缺乏简便的计数系统。当时的计数符合既繁琐又落后，书写和使用都很不方便，只能用来做简单的记录，一旦数目增大，运算复杂，这些原始的符号就尽显弊端了。而没有简便的计数符号，进行概率计算将是十分困难的事，因此计数符号是否先进也在一定程度上决定着概率论的形成。

对于这一点，现代人可能不容易体会得到，究竟古代的计数符号复杂到什么程度呢?文明可以以古罗马的计数系统为例来说明。

古罗马的计数系统是一种现在最为人们熟悉的简单分群数系，大约形成于纪元前后。罗马人创造了一种由 7 个基本符号组成的 5 进与 10 进的混合进制记数法，即

I	V	X	L	C	D	M
1	5	10	50	100	500	1 000

在表示其它数时采取符号重复的方法，如Ⅲ表示 3，XX 表示 20，CC 表示 200 等。但如果数较大表示起来就相当复杂了，比如：

1 999=MDCCCCLXXXXVⅢ

后来为了简化这种复杂的表示法，罗马人又引进了减法原则，即在一个较大的单位前放一个较小单位表示两者之差，如Ⅳ表示 4，CM 表示 900，则

1 999=MCMXCIX.

如果要计算 235×4=940，现代的竖式是

$$\begin{array}{r} 235 \\ \times 4 \\ \hline 940 \end{array}$$

而用古罗马数演算相当复杂，对于这样一个既不含分数和小数，数又很简单的乘法运算处理起来尚且如此复杂，可以想象，即使数学家又足够的时间和耐心，要解决概率计算里涉及的大量纯数运算也是一件太耗费精力的事情。在这种情况下想要做出成果，数学家们的事件不是用来研究理论而只能是忙于应付这些繁重的计算工作了。而事实上，欧洲的代数学相比几何学而言迟迟没能发展起来，很大程度上也是由于受到这种落后的计数系统的限制。不仅仅是古罗马数字，在人类文明史上出现过的其它几种计数系统（如古埃及、古巴比伦等的计数系统）也由于符号过于复杂，同样不能承担进行大量计算的任务。

相反，以位置制为基本原理的阿拉伯数字则比古罗马数字以及古代其它的计数系统要先进的多，它不但书写简便，而且非常有利于加法、乘法的运算以及小数和分数的表示。从上面的例子可以看出，它的使用可以大大节省运算时间，提高运算效率。正是由于使用了这种先进的计数符号，阿拉伯数字的发明者——古印度人的组合数学（组合数学原理是概率计算运用较多的一种数学工具）才得以领先欧洲人许多。据记载，印度人，特别是公元前 300 年左右的耆那数学家就由于宗教原因开展了对排列和组合的研究。公元 400 年，印度人就已经掌握了抽样与骰子之间的关系（比欧洲人早了一千二百年）。而直到公元 8 世纪时，商业活动和战争才将先进的数字符号带到了西班牙，这些符号又经过了八百年的演化，终于在 16 世纪定型为今天的样子了。

数字符号的简单与否对概率论究竟有什么样的影响，我们不妨举例说明。

有 n 个人，当 n 为多少时，至少有两人的生日相同的概率大于二分之一？

假设所有人的生日均不相同的概率为 P，则

$$P=\frac{365}{365}\times\frac{364}{365}\times\cdots\frac{365-n+1}{365},$$

而题中所求概率为

$$P(n)=1-P=1-\frac{365}{365}\times\frac{364}{365}\times\cdots\frac{365-n+1}{365}.$$

通过计算，当 $n=23$ 时，$P(n)=0.51>0.5$，因此答案为 23。

这是概率论中著名的“生日问题”，也是一种很典型的概率计算问题。从它的计算过程中我们不难看出，数运算在概率论中占有重要所地位。如果使用古罗马的计数法，这样一个概率问题从表达到计算都会相当繁琐，以至于它的求解几乎是不可能的。

对于阿拉伯数字的伟大功绩，大数学家拉普拉斯有如下评价：“用不多的记号表示全

部的数的思想，赋予它的除了形式上的意义外，还有位置上的意义。它是如此绝妙非常，正是由于这种简易难以估量……我们显然看出其引进之多么不易”。阿拉伯数字的出现为概率的表达和计算扫清了阻碍，如果没有这些简便的符号，概率论可能还停留在概率思想的阶段。正是由于使用了可以简洁地表示分数和小数的阿拉伯数字，才将概率思想得以通过形式化的符号清晰地表现出来并逐渐形成理论体系。在概率论的孕育阶段，这种形式化的过程是十分必要的，它使得对概率的理解和计算成为可能，因此先进的计数系统对概率论的形成和发展都起着重要的作用。

除了需要具备上述因素外，概率论的形成还需要具备归纳思维。概率论是一门具有明显二重性的理论体系：“一方面它反映了从大量机遇现象中抽象出来的稳定的规律性；另一方面它关系着人们对证明命题的证据或方法的相信程度”。这两个方面都以归纳法作为最基本的研究方法，因此可以说，归纳法是概率论的方法论基础，概率论的产生必须在归纳法被广泛运用的前提下才成为可能。

归纳法虽然是与演绎法同时存在的逻辑方法，但在文艺复兴前，占主导地位的推理方式是演绎思想（不具有扩展性），归纳思维是不受重视的。直到文艺复兴运动以后，这种状态才被打破。归纳法因其具有扩展性而逐渐成为进行科学发现的主导方法。

从演绎到归纳，这个过程实际上是一种思维方式的转变过程，虽然转变是在潜移默化中完成的，但转变本身对于概率论的出现却起着决定性的作用。我们可以通过考察“概率论”（probability）一词的词根“可能的”（probable）来说明这种转变。在古希腊，“probable”并不是今天的这个含义，它曾意味着“可靠的”或“可取的”，比如说一位医生是“probable”就是指这位医生是可以信赖的。但是到了中世纪，这个词的含义发生了变化，它已经和权威联系在一起了。当时的人们在判断事情的时候不是依靠思考或证据而是盲目地相信权威，相信更早的先人所说的话。在这种情况下，如果说某个命题或某个事件是“probable”，就是说它可以被权威的学者或《圣经》之类的权威著作所证明。而经过了文艺复兴之后，人们意识到对自然界进行探索（而不是崇拜权威）才是最有价值的事，正如伽利略所说的那样：“当我们得到自然界的意志时，权威是没有意义的。”因此，“probable”才逐渐与权威脱离了关系。15、16 世纪时它已经具有今天的含义“可能的”，不过这种可能性不再是权威而是基于人们对自然界的认识基础上的。

“probable”一词的演化体现了人们认识事物方式的转变过程。当然这并不是说，文艺复兴以前没有归纳思维。当一个人看到天黑他会自然想到太阳落山了，因为每天太阳落山后天都会黑。这种归纳的能力是与生俱来的，即使中世纪的人们思想受到了禁锢，这种能力还不至于消失。而抛弃了权威的人们比先辈们的进步之处在于，他们是用归纳法（而不是演绎法）来研究自然界和社会现象的。他们将各种现象当作是自然或社会的“特征”，进而把特征看作是某种更深层的内存原因的外在表现。通过使用归纳推理进行研究，他们就可以发现这些内在原因，从而达到揭开自然界奥秘和了解社会运行规律的目的。于是在好奇心的驱使下，归纳思维被充分地激发出来。而这一点恰恰是概率论得以实现的必要条件。从概率论的第一重特性中可以看出，概率论所研究的对象是大量的随机现象，如赌博游戏中掷骰子的点数，城市人口的产生和死亡人数等。这些多数来自于人们社会活动的记录都为概率论进行统计研究提供了必须的数据资料。虽然这些记录的收集与整理其目的并不在于发现什么规律，但善于运用归纳思维的人却能从中挖掘出有价值的烟具素材。例如，

早在16世纪，意大利数学家卡尔达诺就在频繁的赌博过程中发现了骰子的某些规律性并在《机遇游戏的学说》一书中加以阐述；17世纪，英国商人J.格龙特通过对定期公布的伦敦居民死亡公告的分析研究，发现了死亡率呈现的某种规律性；莱布尼兹在对法律案件进行研究时也注意到某个地区的犯罪率在一定时期内趋向于一致性。如果没有很好的归纳分析能力，想从大量复杂的数据中抽象出规律是不可能的。而事实上，在17世纪60年代左右，归纳法作为一种研究方法已经深入人心，多数科学家和社会学家都在不自觉地使用归纳的推理方法分析统计数据。除了上述两人（格龙特和莱布尼兹）外，统计工作还吸引了如惠更斯、伯努利、哈雷等一大批优秀学者。正是由于许多人都具备了运用归纳法进行推理的能力，才能够把各自领域中看似毫无秩序的资料有目的地进行整理和提炼，并得到极为相似的结论：随即现象并不是无穷无规律的，大量的随即现象的集合往往表现出某种稳定的规律性。概率论的统计规律性正是在这种情况下被发现的。

概率论的第二重特性同样离不开归纳法的使用。既然概率论反映的是人们对证明命题的证据的相信程度（即置信度），那么首先应该知道证据是什么，证据从何而来。事实上，证据的获得就是依靠归纳法来实现的。在对自然界特征的认识达到一定程度的情况下，人们会根据现有的资料做出一些推理，这个推理的过程本身就是归纳的过程。当假设被提出之后，所有可以对其合理性提供支持的材料就成了证据，即证据首先相对于假设而言的。如果没有归纳法的使用，证据也就不存在了。由于归纳推理在前提为真的情况下不能确保结论必然为真，因此证据对假设的支持度总是有限的。在这种情况下，使用归纳推理的道德命题的合理性便不能得到充分的保障。而概率论的第二重特性就是针对这个问题的，证据究竟在多大程度上能够为假设提供支持?这些证据本身的可信度有多少?为解决归纳问题而形成的概率理论对后来的自然科学何逻辑学的发展都起到了重要的作用。

归纳法的使用为概率论的形成提供了方法论基础。它一方面使得概率的统计规律性得以被发现，另一方面，也使概率论本身具有了方法论意义。从时间上看，概率论正是在归纳法被普遍运用的年代开始萌芽的。因此，作为一种具有扩展性的研究方法，归纳法为概率论的诞生提供了坚实的思维保障何方法论保障，在概率论的形成过程中，这种保障具有不容忽视的地位。

与前面述及的几点因素相比，社会因素显然不能作为概率论产生的内在因素，而只能被当作是一种外在因素。但从概率论发展的过程来看，作为一种与实际生活紧密相关的学科，其理论体系在相当大的程度商是基于对社会何经济问题的研究而形成的，因此对实际问题的解决始终是概率理论形成的一种外在动力。在这点上，社会因素与概率理论形成了一种互动的关系，它们需要彼此相结合才能得到各自的良好发展。从17、18世纪概率论的初期阶段来看，社会经济的需求对概率论的促进作用是相当巨大的。

在社会需求中，最主要的是来自保险业的需求。保险业早在奴隶社会便已有雏形，古埃及、古巴比伦、古代中国都曾出现过集体交纳税金以应付突发事件的情形。到了14世纪，随着海上贸易的迅速发展，在各主要海上贸易国先后形成了海上保险这种最早的保险形式。其后，火灾保险、人寿保险也相继诞生。各种保险虽形式各异，但原理相同，都是靠收取保金来分担风险的。以海上保险为例，经营海上贸易的船主向保险机构（保险公司）交纳一笔投保金，若货船安全到达目的地，则投保金归保险机构所有；若途中货船遭遇意外而使船主蒙受损失，则由保险机构工具损失情况予以船主相应的赔偿。这样做的目的是

为了将海上贸易的巨大风险转由两方（即船主与保险公司）共同承担。从这个过程中可以看出，对保险公司而言只要船只不出事，那么盈利将是肯定的；对船主而言，即使船只初始，也可以不必由自己承担全部损失。

从性质上看，从事这种事业实际上就是一种赌博行为，两方面都面临巨大风险。而这种涉及不确定因素的随机事件恰恰属于概率论的研究范围。由于保险业是一项于双方都有利的事业，因此在 16、17 世纪得到了快速的发展，欧洲各主要的海上贸易国如英国、法国、意大利等都纷纷成立保险公司，以支持海上贸易的发展。此外还出现了专门为他人解决商业中利率问题的“精算师”。不过在保险业刚起步的时候，并没有合理的概率理论为保金的制定提供指导，最初确定投保金和赔偿金全凭经验，因此曾经出现过很长时间的混乱局面。而这样做的直接后果就是不可避免地导致经济损失。例如在 17 世纪，养老金的计算就是一个焦点问题。荷兰是当时欧洲最著名的养老胜地和避难场所，但其养老金的计算却极为糟糕，以至政府连年亏损。这种状况一直持续到 18 世纪，概率理论有了相当的发展，而统计工作也日渐完善之后，情况才有所改观。在结合大量统计数据的前提下，运用概率理论进行分析和计算，由此得到的结果才更有可能保证投资者的经济利益。

我们可以举一个人寿保险的例子来说明概率理论是如何应用到保险事业中来的：2500 个同年龄段的人参加人寿保险，每人每年第 1 月交投保费 12 元。如果投保人当年死亡，则其家属可获赔 2000 元。假设参加投保的人的死亡率为 0.002，那么保险公司赔本的概率是多少？

从直观上看，如果当年的死亡人数不超过 15 人，则保险公司肯定获利，反之，则赔本。不过单凭经验是绝对不行的，必需有一套合理的理论来帮助处理此类问题。工具所给条件，每年的投保费总收入为 $2500\times12=30\,000$（元），当死亡人数 $n\geq15$ 时则不能盈利。令所求之概率为 P，由二项分布的计算公式可以得出 $P(n\geq15)=0.00\ 0069$。也就是说，如果按照以上条件进行投保并且不出现特别重大的意外，则保险公司有几乎百分之百的可能性会盈利。

这个问题就是通过将概率理论运用到关于人口死亡的统计结果之上从而得到解决的。这个简单的例子告诉我们，概率理论对保险业的发展有着相当重要的指导作用。工具统计结果来确定在什么条件下保险公司才能盈利时概率理论对保险业最主要的贡献，它可以计算出一项保险业务在具备哪些条件的情况下会使保险公司获得收益，并进而保证保险公司的经营活动进入良性循环的轨道。从另一个方面看，最初保险业的快速发展与其不具有基本的理论依据是极不协调的，这很容易导致保险公司由于决策失误而蒙受经济损失。由此保险事业迫切需要有合理的数学理论作为指导。在当时的社会环境下，由科学家参与解决实际问题是非常有效的，而由保险所产生的实际问题确实曾吸引了当时众多优秀数学家的目光。在 1700~1800 年间，包括欧拉、伯努利兄弟、棣莫弗、高斯等在内的许多著名学者都曾对保险问题进行过研究，这些研究的成果极大地充实了概率理论本身。

可以说，经济因素和概率理论在彼此结合的过程中形成了良好的互动关系，一方面数学家们可以运用已有的理论解决现实问题。另一方面，新问题的出现大大刺激了新理论的诞生。概率论的应用为保险业的合理化、规范化提供了保证，正是由于有了概率论作理论指导，保险业的发展才能够步入正轨。反过来，保险业所出现的新的实际问题，也在客观上促进了概率论的进一步完善。这样，对于概率论的发展来说，保险业的需求便顺理成章

地成为一个巨大的动力。

总之，概率论的产生就像它的理论那样是一种大量偶然因素结合作用下的必然结果。首先，赌博这种机遇游戏提供了一种良好的独立随机过程，在进行赌博的过程中，最原始的概率思想被激发出来了；其次，先进的计数系统为概率思想的表达扫清了阻碍，也使得这些思想得以形式化并形成系统的理论。当然在获得概率思想的过程中，思维方式的转变和研究方法的进步才是最根本的关键性条件。如果没有归纳法的使用，即使存在良好的独立随机过程也不可能使人们认识到大量统计数据中所隐藏着的规律性。此外，社会经济的发展，需要借助数学工具解决许多类似保险金的计算这样的实际问题，而这些吸引了众多优秀的数学家们的兴趣的问题对于概率论的形成是功不可没的，它大大地刺激了概率理论的发展，使概率论的理论体系得到了极大的完善。从概率论的历史来看，这几种因素的结合点是 17 世纪末至 18 世纪初，因此概率论在这个时间诞生是很自然的事了。

7.2 实践——认识——实践是概率论历史发展的必然规律

在相当长的时间里，人们认为概率论起源于对赌博的研究，这种观点只看到个别的历史事实和表面现象，而没有看到推动概率论形成的本质因素，这至少是不全面的。虽然在概率论的早期文献中有过一些赌博问题（如帕斯卡解决的由赌徒德·梅雷提出的“点数问题”），但赌博毕竟是少数人的灰暗行为，它不可能成为推动一个学科发展的动力。推动概率论发展的强大动力只能是社会实践的需要。概率论的发展遵循着“实践——认识——实践”的规律。

1.概率论起源于社会实践

17 世纪欧洲文艺复兴时期，意大利的一些城市和中欧的纽伦堡、维也纳、布拉格都相继发展成为新兴的工业城市，随着商业贸易日新月异，于是商业保险，海上运输保险，人寿保险及水灾、火灾保险事业也就应运而生，经济的发展也要求对国民经济收入、税率、人口出生率、死亡率等作出定量的描述、分析和预测，这些领域都向数学提出了新的要求，需要运用数学工具来研究偶然现象中蕴藏着的客观规律，估计事故发生的可能性的大小，这就为概率论的诞生创造了条件。不过，在 1713 年以前，研究随机现象统计规律还处于局限在解决一些具体问题的阶段。

到了 1713 年，瑞士雅各布·伯努利的《猜度术》出版，它第一次陈述了一个带普遍性的定理——大数定律，结束了概率论只限于“单个问题，具体解决”的局面，从而形成了一般的理论，使概率论成为一门独立的数学分支学科。

2.社会实践丰富了概率论的理论系统

1713 年以后，在社会实践的推动下，概率论作为一门充满活力的新兴学科在开拓中前进。

自 18 世纪到 19 世纪初叶，概率论的理论逐渐丰富，此间对概率论的发展有重要贡献的，除了雅各布·伯努利外，还有棣莫弗、贝叶斯和拉普拉斯。

棣莫弗于 1718 年出版了《机遇论》，叙述了一般的概率乘法公式，首次提出了中心极限定理和正态逼近法。贝叶斯在 1763 年初发表了后验概率公式，由它可根据观察结果推断出导致这一结果的未知原因的概率；蒲丰在 1777 年发表了《偶然性的算术尝试》，

把概率论与几何结合起来，开始了几何概率的研究，其中著名的蒲丰投针问题是现在蓬勃发展的概率计算方法的先导。

1812 年，拉普拉斯出版了《分析概率论》，他率先广泛地并始终一贯地将分析工具应用于概率论，因此从本质上说，这就使概率论成为数学有机体中的一个分支，从而增强了它的威力和灵活性。拉普拉斯的另一大功绩是把母函数方法引入了概率论，随意地使用后来称之为拉普拉斯变换的工具，只是由于柯西（Caucky）和阿贝尔（Abel）在收敛问题上的过分小心谨慎，才损害了母函数的名声。

总之，几个世纪以后，概率论发生了质的变化，它研究的对象更加一般，更加抽象，特别是关于中心极限定理的研究在长达两个世纪的岁月里，成为概率理论界最热衷的课题，得出了一系列优秀的结果。

3.概率论广泛地运用于实践

概率论的发展一开始就是与保险事业、人口普查及经济发展方面的应用紧密联系在一起的，随着概率论理论体系的完善，它的应用范围更加广泛，沟通与其他学科联系的边缘学科不断产生，形成了概率论应用的轰动效应。

早在 19 世纪初，高斯与拉普拉斯就将概率论用于物理与天文学数据的误差研究，建立了一般的“误差理论”，引进了最小二乘法。

在 19 世纪和 20 世纪前期，概率论应用得最广泛的一个领域仍是天体力学。1857 年，马克斯韦尔（Maxwell）研究气体运动论，指出了土星环的稳定性；1870 年奥地利玻尔兹曼（Boltzman）以集合论为基础研究了统计力学理论的基础—遍历性理论；1837~1838 年，维纳也着手研究了扰动，把土星环和飞机这两个迥然不同的对象用现代概率论联系了起来。

概率论的另一应用领域是数理统计，特别是生物统计。数理统计的蓬勃发展使概率论的应用深入到物理、化学、生物等实验科学以及工业、农业、经济及社会科学各个领域。

最早建立生物统计思想的是比利时的奎特勒（A.Queelet）。他把统计学的理论应用于解决生物学、医学和社会的问题，引进了“平均人”的概念，第一个认识到大量变异数据中蕴藏着规律性，这正是近代生物学中最重要的思想。1829 年，他分析了比利时的人口调查资料，注意到年龄、性别、季节、职业及经济状态与死亡率的关系，首先作出了自然人口调查统计分类。这种分析对人身保险是有力的支持。最早应用“生物统计”方法的另一位生物学家是赫赫有名的 C.达尔文（Dalwen）（英）。他创立的进化论之实质正是生物统计的规律性反映，他的同代人—奥地利布鲁恩修道院院长孟德尔（Mondel ）于 1865~1866 年间发表了他在杂交上的出色的科学实验成果《植物杂交试验》，破天荒地第一次把遗传和数学联系起来（虽然他的著作在他死后 30、40 年，到 1900 年才被人们重新认识到其重大的价值）。生物统计的飞速发展是在 19 世纪之后，具有特殊天才的高尔顿（E.Galton）（达尔文的表弟）综合了奎特勒和孟德尔的智慧和成就，他在相关性的研究中取得了关于生物数学的又一划时代的成果，创立了回归分析方法和相关系数的检验。

高尔顿的学生威尔登（W.Weldn）研究了独立系统中和在局部种群中的变异，从高尔登的工作中迈出了决定性的一步。1901 年，他和在数理统计学萌芽时期作出了突出贡献的学者皮尔逊（K.Pearson）一道创办了《生物统计（BIOMETRIDA ）》杂志，从而使数理统计在数学发展史上占有了自己的一席之地，其间，他们还开始提出了各种领域内的频数

分布，尤其是皮尔逊还进一步发展了用数学曲线来拟合观察结果的新理论体系。

1908年，皮尔逊的学生戈赛特（W.Gosset）在《生物统计学》杂志上以“STUDENT”（学生氏）为笔名发表了关于t－统计量的精确分布形式，这是统计量精确分布理论中一系列重要结果的开端。

皮尔逊的继承人费歇尔（R.A.Fisher）对数理统计学的发展作出了决定性的贡献。在1912年，他创立了极大似然估计法。1920~1930年他奠定了试验设计的基础，发展了关于多因素试验和区组试验中一系列带根本性的重要概念和方法。

从1928年开始，奈曼（Neyman）在和皮尔逊合作的一系列工作中，建立了假设检验的严格理论，把检验问题明确地归结为数学上一个最优化解的问题。在第二次世界大战中，沃尔德（A。Wald）的《序贯分析》提出了复式抽样方案，系统地发展了“序贯比”检验方法，使二次世界大战中，美国和英国的一些新式武器的试验取得了突破性的进展。

4.概率论的应用促进了理论的深化

概率论应用的范围越广，人们就越是发现它的基础不牢固。古典等可能的概率定义远不能满足实际的需要，于是本世纪初，许多人便致力于改进概率的定义，米赛斯和费歇尔（R.A.Fisher）发展了概率的统计观点，米赛斯还引进了样本空间的概念，从而使概率的严格数学理论建立在测度论的基础上成为可能。在20世纪20年代中期，在法、俄数学家的影响下，概率论的测度论方法逐渐形成，到1936年，最后由苏联数学家柯尔莫哥洛夫所完成。他在其名著《概率论基础》中建立了概率论的公理化体系，从而为概率论奠定了坚实的理论基础。

过去，古典概率论所研究的主要是事件的概率及有限个随机变量的分布，而今现代概率论则主要研究无穷多个随机变量的集合，或简称随机过程。其中研究得较多的有马尔可夫过程、平稳过程、鞅、正态过程、点过程等等。将随机过程研究与其它学科相结合，便产生一些新的边缘学科分支，比如它分别与微分方程、数理统计、数论、几何、计算数学相结合，便产生了随机微分方程、过程统计、数论中的概率方法，几何概率与计算概率等分支；在研究方向上，除了历史上较悠久的极限理论外，还出现了主要由法国学派艾迈尔·波尼耳（Borel）和鲍尔·勃威（Puwal）开创的随机过程的一般理论，轶的现代理论，形成了一个独特的概率理论系统；来源于统计力学的，由无穷多个质点所构成的随机场的理论，点过程的现代理论，马尔可夫过程与位势论等。在中国，许宝禄、钟开莱的工作也是本世纪数理统计学发展的重要组成部分，特别是许宝禄在统计推断和多元分析上都有杰出的贡献。他的工作往往是大量文献的起点，被人们称作“数学严格性”的一个范本。许宝禄以精湛的数学技巧闻名于数理统计界。他推进了矩阵论在统计理论中的应用，而且得到了许多矩阵论的新结果。

由于偶然性几乎无处不在，因此现代概率论的应用也几乎伸展到经济建设、自然科学与社会科学的每个领域；统计物理、量子力学离不开概率论的思想，人们发现，偶然现象在微观世界中比在宏观世界中更为普遍；生物学中研究遗传、群体增长、疾病传染；化学中的反应动力学，高分子的统计性质；天文学中研究银河亮度起伏以及星系空间结构等问题都需要概率论作工具或提供随机模型。在核反应堆中人们利用随机模型来研究中子的减速过程，军事运筹学中的随机搜索与射击模拟等现代自动控制技术需要考虑随机的干扰，这就需要利用随机微分方程来描述状态的转移，因此，概率论成为控制论必不可少的工具。

在通讯技术中为了排除随机噪声而发展起来的滤波理论和一般的数学信息论都离不开概率论。至于在气象、地震、病虫害及人口预测预报、实验设计、质量控制、抽样检查以及经济数学、排队论、运筹学中广泛使用的概率方法，则更是久经考验和行之有效了。

7.3　随机性的认识论

随机数学已经发展了 300 多年，但“随机性究竟是客观存在还是由于我们认识能力的局限性导致的对事情的结果无法确切预知”？这一自古希腊时就已经开始探讨的哲学问题，至今仍然耐人寻味。

1.古河流文明时期，随机性是神的意愿的表达

考古证据显示在早期文明中，人们就已经开始制造和使用多种多样的随机函数发生器做一些严肃的决策或纯粹进行机会游戏。事实上，大约在公元前 3000 年左右，古河流文明地区的人们——古巴比伦、古印度和古埃及人就已经用黏土制作三棱柱、四棱柱和正六面体等不同形状的骰子，并使用象牙赌板、动物趾骨、坚果核等各类随机函数发生器。在中国，考古发现的甲骨卜辞和数字卦显示在殷商时代龟卜成象等活动就已很盛。注意到中国的第一个计数系统是甲骨文计数系统，约形成于公元前 1600 年；古巴比伦的楔形数字系统出现于公元前 2400 年左右；古埃及的象形文字系统形成于公元前 3500 年左右；虽然印度的达罗毗荼人的象形文字至今不能解读，但由吠陀时期印度的数学状况以及与其它文明的类比不难断定那时印度应该是已经有了记数系统的，因此作者认为，原始赌具的发明年代与各文明古国记数系统的产生年代相当；因此人们应当在更早的时候就对随机性有所意识了。

也许是由于原始社会的生产力极为低下，原始人类在千变万化的自然现象面前惊恐不安，从而产生了万物皆有灵的自然观并试图找出沟通神灵的途径，从而产生了原始宗教和祭司、占卜等活动。由一些中外古文化遗产不难看到：在这些活动中，人们有一种强烈的信仰，即结果是由神灵操纵的，而签或骰子之类的随机函数发生器则是体现神灵意愿的圣物；为了尽可能消除人类的影响、准确再限神谕，人们往往有一套旨在消除认为影响的程序例如闭上眼睛或让通常被认为没有自我意识的幼儿抽签或抛掷骰子等。

总之，在早期文明中，人们就对或然性有所认识。当时人们普遍认为随机性表达的是神灵的意愿；各文明古国人们有意识地利用随机性的时间应当不晚于其记数系统的产生年代。

2.古希腊时期，随机性是尚未认识的确定性与随机性属于上帝和自由意志的对决

随着古埃及、古巴比伦被征服，这些古国灿烂的文明向西方传播并被发展成更加理性的古希腊海洋文明。古希腊人继续利用随机性作决策和进行游戏；古希腊哲学家则开始思考“随机性究竟是客观存在的呢，还是由于我们认识能力的局限性而导致的对事情结果无法确切预知”？对此，古希腊人的观点大致可分为两派，即以原子学派为代表的否定随机性客观存在的一派和以亚里士多德为代表的认为随机性客观存在的一派。

古希腊人认为：自然是有序的，按完美的设计恒定地运行着。毕达哥拉斯学派认为“序”就是数，他们的世界观是“万物皆数”。毕达哥拉斯之后的哲学家们更加关注现实世界的本质和基本的数学设计。以路希帕司和德谟克里特为代表的原子学派坚信一切都是有原因的，随机性并不意味着没有原因而是意味着隐藏的原因。 路希帕司说：“没有什么事物

是随机发生的；每一事物的发生都有其原因和必然性”；其继承者德谟克里特继续坚持这个观点认为或然性意味着对确定性原因的无知。亚里士多德是第一个承认并尝试解释随机性的古希腊学者。他将事件分成3类：

（1）必然发生的确定事件。

（2）多数情况下要发生的可能事件。

（3）发生与否完全随机的不可预知事件。

亚里士多德认为第三类事件是世界上最脆弱的事情，是偶然引起即碰巧发生的，其原因不能确定。亚里士多德后的希腊哲学家们依然对随机性保持哲学上的兴趣，并进行着长期的争论。如后期斯多葛派学者们继承了路希帕司的观点，认为“没有无原因的事件或自发性的行为”；而伊壁鸠鲁学派则坚持认为不确定性是客观存在的。

众所周知，古希腊人在数学上取得了巨大的成功，但是，没有任何迹象表明他们曾经计算过随机事件的概率。事实上，直到15世纪末16世纪初才开始有一些意大利数学家计算概率；而这一学科的真正建立，则要等到17世纪中叶。综观这漫长的历史进程，人们不禁深感困惑：既然人类在很早的时候就已经意识到并反复利用或然性，为什么这个领域的发展却如此缓慢——尤其是，古希腊人为什么没利用机会游戏的对称性和频率稳定性建立与欧式几何结构颇为相似的古典概率?对此，一些学者进行了讨论，归纳起来主要有以下三点：首先是亚里斯多德的哲学观限制了古希腊的科学家们。因为亚里斯多德关于事件的分类广泛流传并被普遍认同，因此人们一般认为机会游戏的结果属于第（3）类事件，不可能通过科学研究来理解。其次，古希腊人所使用的随机函数发生器——趾骨并不完全对称，因而没有等可能选择的经验例子，所以古希腊人未能认识到古典概型的对称美。另外，宗教原因也阻碍了人们对偶然性进行研究——不可预知的随机事件是由上帝预先决定的，而对或然性进行研究无疑是插手神灵的决定。我认为，亚里斯多德的哲学观和宗教原因限制了古希腊数学家研究随即现象的数量规律的说法是有道理的；但认为当时没有等可能选择的经验例子，从而阻碍了他们对概率进行研究的观点却是站不住脚的。因为正像我们在前面所说的，早在公元前3000年左右古河流文明的人们即已发明了非常对称的骰子，因此没有理由认为聪明的古希腊人只用动物趾骨作为随机函数发生器；况且古希腊人通过摇签作决定是很普遍的，而这是一种非常典型的等可能概型。

综上所述，在一千多年的时间里，古希腊人对随即现象的认识大致分为两派，但亚里斯多德的哲学观和宗教原因阻碍了他们对随机现象的数学研究，使他们对随机性的讨论只停留在哲学层面。

3.中世纪及近代数学时期，随机性不存在

在中世纪的欧洲，宗教禁锢着人们的思想，决定论的世界观占据绝对优势，人们普遍认为万事万物不论大小都是在上帝的指引下发生的，因此这时对随机现象是不可能有进一步研究的。15世纪末、16世纪初，文艺复兴的意大利挣脱了中世纪严酷的宗教禁锢，经济、文化、政治一派繁荣；同时，赌博开始盛行、保险业正在兴起、彩票发行日趋普遍、社会及人口统计达到较高水平。这些游戏和行业的客观需要导致了概率论的早期探索。帕乔利、塔塔利亚、卡尔达诺、伽利略等率先讨论以骰子为代表的机遇性赌博中各种具体情况出现的可能性的计算问题。

17世纪中叶开始，人们对随机现象的研究进入了一个崭新的阶段。据文献记载，在

1646 年前后，英国哲学家霍布斯与伦敦德里主教拉姆霍尔博士就此展开了一场争论。在这场争论中，霍布斯坚持认为，所有事件都是有上帝或上帝决定的外在原因预定的，而正是对那些原因的无知，使一些人认识不到事件的必然性，因而把它们归属于或然性。而拉姆霍尔则以掷骰子为例指出：或然性结构的物理过程一经发生便由物理规律决定因而是决定性的，但这种结构有如此之多的不确定方面以至于在发生之前是非决定性的。在这场争论后 8 年概率论诞生了，其标志是 1654 年帕斯卡与费马的 7 封通信即著名的“点数问题”。从此，数学家们开始密切关注和深入研究随机现象。

需要指出的是，虽然这时研究随机现象的理论诞生了，但是人们，包括概率论的奠基者们似乎都普遍认为随机性其实并非客观存在，承认它不过是因为无知而不得不采取的一种权宜之计。例如 1692 年出版的惠更斯的《论赌博中的计算》英译本的前言中写到：“具有这样一个决定性的力和方向的一枚骰子，不落在这样一个决定性的面上是不可能的，只是我不知道使它落在这样一个决定性的面上的力和方向，所以我称之为或然性，那不是别的，只是技术上的需要。”又如，这一时期概率论的集大成者法国数学家拉普拉斯 1814 年在他的《概率的哲学导论》中写到：“我们应该把宇宙的目前状态看做是它的先前状态的结果，并且是以后状态的原因。万能的指挥者能够在某一瞬间理解使自然界生机盎然的全部自然力，而且能够理解构成自然的存在的各自的状态，如果这个智慧者广大无边到足以将所有这些资料加以分析，将宇宙中最巨大天体的运动和最轻的原子的运动都包含在一个公式中。那么对于这个智慧者来说没有任何事物是不确定的，未来如同过去一样在他的眼中将一览无余。人类的心志在致力于天文学上所表现出来的完美中，给出了与这一智慧者微弱的相似性。人类在力学和几何学上的发现，加上万有引力的发现，使它能够用相同的分析表达式去理解宇宙系统的过去状态和未来状态。通过把同一方法应用于某些其他的知识对象，人类已能将观察到的现象归结为一般规律，并且预见到给定条件下应当产生的结果。所有这些探索真理的努力都倾向于引导人类心智不断地回到我们刚才所提到的广大无边的智慧者那里去，但是，人的心智距这一智慧者仍有无穷之远。”

有人认为，之所以出现这种状况，根本原因是 17 世纪牛顿经典力学的建立。牛顿力学的胜利使人们认识到自然界因果联系的客观性和普遍性，并形成了认识论上的决定论和可知论，即把精密性、序、法则、涉及和必然性视为可以被认识的责任的本质，而把随机性等同于无知。在此后 200 多年的事件里，科学家始终坚信世界是上帝用数学精确设计的，因此也是可以用数学破译的：一切事物的变化都遵从确定的规律，即使考虑到偶然的因素，只要近似地知道了一个系统的初始条件和变化规律，就可以计算出它的近似行为；随机性并不存在，对结果的不确定仅仅是我们无知的表现，我们知道的越多，不确定性就越少；倘若我们知道得足够多，至少在理论上我们就能消除所有的不确定性。正是在这样的认识论基础上，科学家们以大无畏的精神探索着许多以往被认为是非常困难的科学——包括概率论，并取得了辉煌的成就。

总之，这一时期，人们普遍认为随机性并非客观存在、承认它不过是权宜之计，所以有人认为这是认识论上的大倒退。但我认为正是这种认识促进了对随机现象的数学研究，因为数学家们正是在破解“世界的数学精确设计”的坚定信念下大大发展了概率论：不但组合概率论在这一时期得到了大发展，而且微积分被成功地应用于随机现象的研究并形成了分析概率论。

4.现代数学时期，确定性只是随机性模式的一种特殊情况

在概率论诞生后200多年的时间内，数学家们基本上是在决定论的观点下安然地研究概率论。然而在关于误差的研究中隐藏着一股认识变异的潜流：科学家们意识到测量误差是不可避免的——无论测量工具多么精确，无论操作者多么仔细，都无法从根本上消除。世界并非完全确定，确实存在着无法回避的随机性——这些思想在19世纪受到注意，在20世纪被越来越多的人接受。

我们认为，数学理论的发展为随机性认识的转变提供了智力环境；而数学的应用，尤其是在现代物理学上的应用，则为这种转变的实行提供了契机。

影响数学家们对随机性认识的数学事件无疑有很多，但我们认为主要是非欧几何的建立和四元数的出现使数学家们对确定性初始了决定性的动摇。我们知道，16~19世纪上半叶在决定论观点的指导下，科学家们在探索自然设计奥秘的过程中取得了辉煌的成就。虽然“自然是上帝的数学设计”这一信条被数学家们的工作逐渐削弱，人们也对数学家工作的真理性产生了怀疑，但总的来说，直到19世纪初，数学家们仍然相信严格的数学真理和自然界的数学法则。在所有的数学分支中欧式几何最受推崇——两千多年来，它一直保持数学严格性典范的神圣地位，许多数学家相信它是绝对真理。然而，这个近乎“科学圣经”的欧式几何却一直是“白璧微瑕”——那个瑕疵就是平行公设，是欧式几何最终受到致命袭击的“软肋”。19世纪20~30年代，俄国数学家罗巴切夫斯基和匈牙利数学家J.波约以及德国数学家高斯从否定平行公设乘法，分别独立建立了双曲几何———种与欧式几何完全不同、但推理一样无懈可击的几何体系。1854年，德国数学家黎曼又进一步发展了罗巴切夫斯基等人的思想，建立了一种更广泛的几何——黎曼几何，并使双曲几何、欧式几何都成为它的特例。1870年代以后，意大利数学家贝尔特拉米、德国数学家克莱因和法国数学家庞加莱等人先后在欧式空间中给出了非欧几何的主观模型，揭示了非欧几何的现实意义，从而使非欧几何获得了广泛的理解和接受。我认为这不仅对人们的时空观产生了极其深远的影响，而且极大地动摇了人们对于真理与自然法则的根深蒂固的认识，引起了人类认识史上的革命，也使得随机性在数学家们的心中的地位开始发生变化——既然欧式几何都不再是唯一绝对的真理，而只是若干种可能性中 的一个，是一种特例、一种近似，那还有什么是绝对的和确定的呢?

1843年，爱尔兰数学家哈密顿发现了四元数，这是历史上第一次构造出不满足乘法交换律的数系。此后，各种新的超复数如雨后春笋般涌现出来：德国数学家格拉斯曼开始考虑n维线性空间（《线性扩张》，1844）；闵可夫斯基、希尔伯特又分别建立了将时间和空间结合在一起的闵可夫斯基空间（1907）和无穷维空间希尔伯特空间（1904~1910）。我们认为，随着数系和空间概念的一步步推广，数学家们进一步认识到：在数学王国里没有绝对真理和终极真理。

接着，数学物理学家们心目中的确定性也动摇了。十九世纪数学物理学家们很快为上述新的数学工具找到了实际应用：英国数学物理学家麦克斯韦将概率方法引入气体运动的研究，美国数学物理学家吉布斯和英国数学物理学家亥维赛又将四元数发展为物理学中广泛应用的向量代数和向量分析；基于黎曼几何的绝对微分学为广义相对论提供了合适的工具；希尔伯特空间正是量子力学的合适工具；波恩则“在闵可夫斯基的工作中找到了（狭义）相对论数学的整个武器库”；随着工具越来越齐备，物理学家越来越充满信心地将其

研究对象向大和小的两个极端扩展和深入——在这个过程中，他们心目中的确定性动摇了，他们对于随机性的认识改变了。

首先，麦克斯韦1859年首次将概率观点和统计方法引入气体动力学。他认为，每个分子的运动复杂得难以用牛顿定律描述；而气体的性质是整体的性质，考虑组成气体的每个分子的运动远不如考虑整体运动重要；因此，应当通过整体的概率描述而不是每个分子的精确运动来理解气体。我们看到，对麦克斯韦来说，概率论方法是必要的，他基于概率论的气体模型揭示了看待自然的新方式。

其次，对布朗运动的研究引起了科学思维方式的转变。1828年英国植物学家布朗最早注意到水中花粉的不规则运动，即所谓的布朗运动。1905年和1906年，爱因斯坦和波兰物理学家斯莫鲁霍夫斯基分别独立发表论文，以不同方式正确地定量解释了布朗运动，布朗运动才获得重视。人们发现布朗运动本质上是不可逆随机过程，任何非随机性理论都不能解释它。这个结果揭示了自然新的一面：对于不可逆过程而言，随机性不能被忽视。

20世纪20年代，量子力学创立后不久，爱因斯坦与玻尔、波恩等几位主要的量子力学家在认识论的原则问题上产生了严重的意见分歧并由此引发了决定论和非决定论的大论战。论战从1927年的第五届索尔维物理学会议开始，一直延续了几十年，在爱因斯坦和玻尔谢世之后仍未罢休爱因斯坦1944年7月给波恩的信中将这场论战的焦点比喻为“上帝掷不掷骰子”。爱因斯坦属于维护“决定论”传统的一方，他坚信“上帝不掷骰子”，量子统计学只不过是我们长期以来还不能“完整地描述事物”而采用的权宜之计。而以波恩和玻尔为代表的量子力学家们则属于拒绝“决定论”传统的一方，他们认为经典力学中遍及一切的、必然的因果规律以及行之有效的“决定论”模式，是物理世界“深层”概率规律的宏观实现；这种规律和模式只近似地描述宏观物理世界“表层”而不适用于微观物理世界；一旦进入微观世界就必须采用非决定论模式，选择概率、随机性等概念和统计规律。

总之，随着科学研究的深入，科学家们越来越深切地认识到承认随机性不再是无知；概率作为一个基本概念，已不再是我们的无知造成的一种心态，而是“新理性的组成部分”，是自然法则的结果。“贝特朗悖论”，“Maxwell-Boltzmann统计”与现实的距离，Bose-einstein粒子模型、Fermi-Dirac粒子模型的建立以及有关事件的概率的计算进一步揭示了对于随即现象认识的任何绝对真理观的局限性，而概率论公理化体系的建立可以认为是对于这种局限性的有力挑战。所有这些都在向人们表明对于随机性的认识和随即现象的数学研究没有绝对真理和终极真理。

7.4　从概率论的发展看东西方文化传统的差异

西方古典概率思想对概率论这门学科的进一步发展以及西方社会价值观的形成和发展都具有重要的意义，概括起来主要有以下几个方面：

（1）西方古典概率思想不但是一切概率理论和数理统计的基础，而且开辟了从偶然中求必然、从中求确定的新境界。从其产生、发展的过程来看，它的基础作用是毋庸置疑的。

（2）概率问题的背后其实质是一个偶然性中的必然、从不确定中的确定及其度量问题。从赌博到商业冒险，从海外淘金到登越星球，从人口统计到国家选举，无不显示了古典概率思想的广泛应用价值。

（3）“机会均等”或“等可能性”是西方古典概率思想产生或研究的前提。而古典概率论中的所有公式、定理都是此基本思想在数学上的具体体现。“机会均等”的社会人文意义在于在后来的资产阶级文化启蒙运动中，为“机会均等”社会价值观的确立开辟了道路。西方古典概率思想曲折反映了大革命前夕处于上升阶级的资产阶级知识分子对封建制度的不满和对人类自身前途、命运的关切。处于上升阶段的资产阶级知识分子之所以倾注大量精力，研究和认识偶然中的必然性，就在于努力在急剧的社会变革中掌握自己的前途和命运。

从偶然性中求必然、不确定性中求确定的探索精神，尽管从数量上把握有利事件出现的可能性的大小，对机会均等和公平游戏规则的普遍追求，将概率论及概率思想广泛应用到与人类自身命运密切相关的自然、社会研究中去等诸多方面，都是西方概率论及其概率论思想值得称许的地方。

但是我们决不能将古典概率思想的作用夸大到不适当的程度，它有其自身的局限性。我们认为，仅从是否有利和数量上认识把握事件及其出现的机会是西方古典概率论及概率思想的主要局限。这一点连雅各布·伯努利也注意到了。雅各布·伯努利在他的《猜度术》中写到：无论怎样，想要就某事的发展做出一个正确的预测，似乎只需要精确地推算出各种可能出现的结果，然而确定出其中一种情况比另一种情况更易发生的可能性，然而，我们很快又会遇到麻烦，因为这种方法只在很少的一些情形下适用。试问，谁能够告诉所有疾病的种类，推算出所有即将发生的病例，并说出一种病症比另一种更加致病的可能性。或者，谁又能够列举出大气每日所经历的无数变化，并据以在当日预报出此后一个月，甚至一年内的天气?正由于雅各布•伯努利感觉到自己在寻求不确定量的量化测定过程中并不成功，所以他并没有把关于概率方法在政治、伦理、经济上的应用写入书中。

总之，从数量上把握有利机会和有利事件，固然是人情常理和智力兴趣所在，但在大多数情况下，单凭计算是无济于事的。就此而论，美国数学史家克莱因“正是一个事件是否发生的概率，决定了我们对该事件的态度和行动”这句话就值得商酌。事实上，很多情况下，正是一个事件的性质或对我们的利害关系及其发生的概率，决定了我们对该事件的态度和行动。

卡尔达诺和拉普拉斯所处的年代正是西欧从封建社会到资本主义社会的转折时期。为了在急剧的社会变革中掌握自身的命运机会与机遇、偶然性与不确定性、随即现象等问题自然成为资产阶级知识分子研究的对象。拉普拉斯曾经说过：一门开始于研究赌博机会的科学，居然成了人类知识中最重要的学科，这无疑是令人惊讶的事情。但如果我们对概率300年间西方文化传统、社会精神风貌和价值观作管窥，这又是一件不可思议的事情。

卡尔达诺所处的时代，发源于意大利的文艺复兴运动已进行了一百多年，商品经济发达。那时，欧洲人开始冲破基督教和神学的束缚，人们崇尚个性解放、个人奋斗，对财富的追求和商业冒险已成为社会共识。像赌博这类活动，和到欧洲大陆之外探险淘金一样自然，所以他们对赌博的认识不像现在的中国人那样往往与个人的道德品行甚至犯罪相联系。受西方文化和价值观的影响，澳门的赌博被冠以“博彩业”并被作为支柱产业受到行政长官的重视，但若在内地这是不可思议的，而美国的拉斯维加斯更是以赌城闻名于世界。

从广义上讲，赌博和其它具有竞争性的游戏一样并无本质上的区别。但西方人更加注重的是机会是否均等以及游戏规则是否公平，即在保证机会均等和遵守公平的游戏规则的

前提下，人人都有参与游戏的权利和获胜的希望。他们把这一价值观和理念推而广之，逐渐形成现在的法治社会。

古希腊文明是西方文明的源头之一。雅典在公元前 5、6 世纪是一个经济高度繁荣，生产力高度发达，民主制度臻于完善的城邦国家。在这个基础上产生了光辉灿烂的古希腊文化，对后世有深远的影响。可以好不夸张地说，现代的很多文学、艺术、哲学及数学思想都可以从古希腊那里找到萌芽。这里我们特别要提及的是在此时期出现了很多对后世影响很大的哲学家兼数学家。如泰勒斯、毕达哥拉斯、芝诺、苏格拉底、柏拉图及亚里斯多德等。从这些先哲的思想中我们可以看到西方理性传统的起源。他们认为，世界是一个有序的、可预言的世界，事物按其本性在其中运作，要认识事物的运作规律只有借助理性；真正的知识是理性的知识，真理只有借助于理性才能获得。柏拉图认定数学认识是理智，坚信宇宙是按数学规律设计的。柏拉图还认为，只有通过心智活动所认识到的数学规律才能够成功地说明自然现象，他在《裴里布》中指出，每门科学只有当它含有数学时才称其为科学。柏拉图数学化的宇宙观正是近代和现代科学数学化思想的重要渊源。柏拉图还明确提出了应从自明的假设出发进行严格的证明，这就是古希腊数学方法的最高成就——公理化方法的发端。

古希腊的先哲们对宇宙间各种各样、千奇百怪的现象产生了好奇心，于是他们设立各种题目，对之溯本求源，不断地进行深入研究，哲学和数学因而得到了迅速的发展。通过研究我们发现，似乎没有哪个古代文明国家像古希腊哲学家那样对量变质变规律作过深入缜密的思考，也没有哪个古代文明国家有希腊那样成熟、严密的逻辑思维，除古代希腊之外，似乎没有哪个古代文明国家对事物的认识都要求确定、准确、可度量、可计算、可论证、可推理、可证伪，即使连文字，他们也要用抽象、简单的字母组合来表示，抽象、简洁、完美成了古希腊的执著追求。如果我们对世界文明史上的“希腊现象”不表惊叹，那么我们对西方文明之树上结出的“概率论”之果又何必大惊小怪呢?自文艺复兴之后，欧洲人已获得了相对的学术研究自由，当然，他们也付出了像布鲁诺那样惨痛的代价。自伽利略（伽利略主张用实验方法研究自然，用数学方法描述自然）之后，西方实验科学普遍兴起，风气所至，他们善于从对自然科学的探索中寻求新知识和乐趣，因而我们对西方数学家和哲学家用“随机试验”的方法研究概率问题也就不足为奇了。

7.5 数学分析对概率论的渗透与推动

数学分析和概率论数学家族中资历很深的两个成员。两者行世的时间差不多，但发展的路径却大相径庭。据史载，费马于 1629 年，便在名曰《求最大值和最小值的方法》的手稿中，论及了求切线的方法；1654 年，费马在与帕斯卡的通信中，解决了由法国赌徒德· 梅雷提出的“点数问题”，这是古典概率中的一个典型案例。由于费马是解析几何、微积分、概率论的最重要的奠基人之一，据此不难推断，数学分析、概率论的历史，时间上限差不多，都只有 360 年的历史。

微积分创立后，先后引爆了数学史上的第二次、第三次数学危机，数学分析内部的攻伐聚讼，进入 20 世纪后才算基本平息。在概率论的发展过程中，伯努利、棣莫弗、辛普生、蒲丰、拉普拉斯、高斯、泊松、切比雪夫等数学家，作出了积极的贡献，除了蒲丰的

"投针问题"具有浓厚的悖论色彩外，其余的成就都是不存争议的。1933年，前苏联数学家柯尔莫哥洛夫以集合论、测度论为工具，导入了概率论的公理化体系，使概率论中断烂丛碎的理论片段，获得了崭新的分布顺序，首次以一个独立学科的面目，出现于数学领域，成为有别于微积分的后起之秀。

半个多世纪以来，概率论突飞猛进，获得了举世公认的进步，在五花八门的类型研究中，改了来了的学科地位持续跃迁。按照流行的数学学科"三分法"，数学可以分为确定性数学、随机数学、模糊数学三大类，而数学分析、概率论分别是前边二者的典型代表，因此概率论不仅具备了与微积分分庭抗礼的地位，更因其非线性、反因果的非理性特征，显得比经典的数学分析更具有时代精神。

概率论的先锋色彩是不容置疑的，但是曾经哺育和推动过概率论的微积分，并非是落拓背时的代名词。在概率论的间架结构中，因果论、确定性的烙印四处可见。应该说，概率论的大厦，是建筑在微积分的地基上的，而概率论的调色板，则始终是以数学分析为底色的。在此意义下，寻绎数学分析在概率论中的地位，阐述概率论的因果论特征，是很有意义的。

1.集合论与概率论的公理化体系

集合论是在微积分的营养液中培育出的一颗明珠。19世纪末，康托尔的朴素集合论，将第三次数学危机推向高潮。随着康托尔悖论、罗素悖论的出现与廓清，公理集合论应时而生；公理集合论使微积分的纷争彻底休止，声势浩大的数学公理化运动宣告开始。

众所周知，数学的研究对象一般都是内涵着某种结构的集合，或者是可以通过集合来定义的事物，因此说，集合论可以充当整个现代数学的基础，在这点上，数学分析和概率论都不例外。由于集合论与微积分之间，存在着明显的源和流的关系；又由于勒贝格积分有效地建立了集合论与测度论的联系，进而形成了概率论的公理化体系；因而集合论对概率论渗透，可视为微积分对概率论的一次较有力的推动。

数学分析中的积分主要有黎曼积分和勒贝格积分两种。黎曼积分在对付性质良好的函数时得心应手，但在遇到级数、多元函数、积分与极限交换次序等较为棘手的问题时，常常夹脚受窘。勒贝格积分出现后，黎曼积分遇到的难题迎刃而解，微积分随之进化到了实变函数论的新阶段。概率论的公理化，就是凭借勒贝格积分的理论，宣告完成的。

建立概率论公理化体系的倡议，是由希尔伯特于1900年的国际数学家大会上提出的，属于著名的23个问题中的第六种。最先涉足该问题的，有庞加莱、波莱尔、伯恩斯坦等人。法国数学家波莱尔是巴黎高师1893年的毕业生，数学分析中的有限覆盖定理便是波莱尔的杰作，以他的名字命名的波莱尔可测集，构成了测度论的基础，波莱尔率先将概率论与测度论挂起钩来，定义了可数事件集的概率，这项贡献填补了古典有限概率和几何概率之间的裂谷，苏俄数学家伯恩斯坦于1917年构造了概率论的第一个公理化体系；1919年，奥地利数学家米泽斯，将概率的频率定义与统计定义进行了公理化的处理，并创立了概率的频率理论学派；有"布尔什维克主义的死敌"之贬称的英国经济学家J.M.凯恩斯也作出了一定的贡献，凯恩斯将任何命题都看成是事件，将事件的概率看成为人们根据经验对该事件的可信程度，凯恩斯定义的概率与随机试验十分隔膜，因而是一种主观概率。主观概率学派虽然没有为概率的公理化运动作出过实质性的贡献，但却在贝叶斯统计学派那儿找到了新的天地。

主观概率学派与概率的频率理论学派的长期对峙与无所作为，致使概率论的公理化运动改道测度论的河床。事件与集合非常相似，概率与测度一拍即合，显示出了雷同的相互关系及运算规律。不仅如此，随着大数定律研究的不断深入，概率论与测度论的深层联系愈加明晰，测度论中的几乎处处收敛与依测度收敛，实质就是弱大数律与强大数律中的收敛。1933 年，柯尔莫戈洛夫出版了数学名著《概率论基础》，柯氏在书中提出了概率的测度论定义，从而统一了概率古典定义、几何定义及频率定义纷争不一的局面。他建立的公理化体系，具备了独立性、无矛盾性、完备性的公理化特征，确立了事件与集合、概率与测度的关系，为集合论加盟概率论，扫清了障碍，寻找到了用武之地。因此，柯氏的公理化体系一经提出，几乎立即获得了举世公认，为概率论的健康发展，奠定了坚实的基础，概率论的发展史已经证明了这一点。

随机变量和分布函数通常记作随机变量 X 与 $F(x)$，它们是概率论中最为重要的两个概念。随机变量 X 是定义在样本空间 S 上的集函数，随机变量 X 达成了简化概率函数的深化或泛化，$F(x)$ 是较之分布率、概率密度函数更为一般的函数，较之随机变量 X 其应用和性质要复杂得多，涉及多元分布函数时，尤其如此。究其实质来看，随机变量 X 为集函数，$F(x)=P(X\le x)$ 为实函数，后者似乎比随机变量 X 初等，不过究其层次来讲，随机变量 X 是 $F(x)$ 的基础，属于 $F(x)$ 的“初级阶段”。然而不管怎样说，随机变量 X 与 $F(x)$ 都是真正的函数。

函数一词始出于莱布尼兹，但涵义与现代不同；函数的原初定义，于 1718 年由约翰 ·伯努利给出，在此五年前的 1713 年，约翰 · 伯努利的长兄雅各布 · 伯努利的巨著《猜度术》，在瑞士的巴塞尔出版，描述概率的统计定义的合理性的“伯努利大数定律”，就包含在这本《猜度术》之中，这是概率论历史上最早最重要的一个定理。随着时间的推移，伯努利大数定律推广到了最一般的情形，名称计有切比雪夫大数定律、马尔可夫大数定律等等多种，其前提条件亦从独立同分布放松为一般的马尔可夫条件

（1）$D(\sum_{i=1}^{n} X_i)<+\infty$。

（2）$\lim_{n\to+\infty}\frac{1}{n^2}D(\sum_{i=1}^{n} X_i)=0$。

而函数的概念则向着精细、严谨的方向进化，欧拉先明前后，曾给出函数的三种不同的定义，其涵义分别为“解析表达式”、“依赖变化”、“由曲线所确定的关系”，其中最后一种定义最接近函数的定义，因而对后世的影响亦最大；1837 年，狄利克雷给出的函数定义，已经接近现在的通用定义，即：“对于在某区间上的每一个确定的 x 值，y 都有一个或多个确定的值，那么 y 叫做 x 的函数。”康托尔的集合论创立后，函数才被定义为集合间的对应关系：1921 年，波兰数学家库拉托夫斯基以关系式

$$(a,b)=\{\{a\},\{a,b\}\}$$

定义了“序偶”的概念，据此可得一般的函数定义如下：设 f 是一个序偶的集合，如果当 $(x,y)\in f$，且 $(x,z)\in f$ 时，$y=z$，则称 f 为一个函数。该定义目前世界通用。

在函数关系的对应下，随机事件先是被简化为集合，继之被简化为实数，随着样本空间转化为数集，概率相应的由集函数约化为实函数。以函数的观点衡量分布函数 $F(x)$，$F(x)$ 的性质是十分良好的，单调有界，可积，几乎处处连续，几乎处处可导。因之，数

学分析中有关函数的种种思想方法，可以通畅无阻地进入概率论领域。随机变量X的数字特征、概率密度与分布函数的关系，连续型随机变量X的概率计算等等，显然借鉴或搬运了微积分的现成成果。

随机变量X与$F(x)$的导入，从理论上结束了概率的古典时代，概率论从此踏上了蓬勃发展的大道。不难确知，概率论的公理化、体系化的动力源，不仅是集合论和测度论，更重要、更基本的，仍然是数学分析那一套理论。因此，概率论形成体系后的高歌猛进，不妨视作概率论向着微积分的靠拢与回归。这种复归的意向对于数学家来说，也许是无意识的，但更大的可能性是一种下意识的有意为之。因为微积分形世几百载，已经形成了一整套完备的理论武器，完全置之不理，显然已不可能。既然测度论已楔入其中，黎曼积分与勒贝格积分的渗透，就在情理之中了。

在函数的范畴内，随机变量X与$F(x)$的本质是一致的。虽然说随机变量X的导入方式有一定的自由度，不具备唯一性；虽说随机变量X的取值需服从一定的概率分布；虽说$F(x)$可以集函数视之，可以描述任何种类的随机变量X的随即性质，但是它们既然是函数家族的成员，就难逃确定性的宿命，就抹不去因果率的印记，这是显而易见的。

1797年，拉格朗日在《解析函数论》中，指出了凡是函数都能用幂级数表示的事实；1807年，傅立叶向巴黎科学院提交了一篇关于热传导的论文，傅立叶在论文中说："任何函数都可以表示成三角级数。"此篇名曰《热的分析理论》的文章，虽然由于拉格朗日的阻挠未获通过，但"傅立叶级数"却不胫而走，成为数学分析中十分有效的工具。1869~1874年间，康托尔在哈勒大学任教，他研究以2π为周期的函数$f(x)$在$(0,2\pi)$上的三角级数

$$\frac{1}{2}a_0+a_1\cos x+b_1\sin x+a_2\cos 2x+b_2\sin 2x+\cdots \tag{7-1}$$

时，意外地碰到了结构复杂的实数点集，经过一番周折，康托尔引入了点集拓扑的概念，谱出了集合论的新篇章。

狄利克雷第一个给出了函数$f(x)$的傅立叶级数收敛于它的自身的充分条件，他的收敛判别方法，即狄利克雷—若尔当判别法。值得重视的是，狄利克雷关于函数的正确概念，正是来自于傅立叶级数的收敛问题。勒贝格积分出现后，对傅立叶分析产生了深远的影响。

对概率论产生影响的，不光是傅立叶级数，还有等比级数、二项和式、调和级数等，作用是方方面面的，有的构成反例、有的便于计算、有的揭示出了特殊的计算方法，等。

假设离散型随机变量X的分布率为

$$P(X=(-1)^{j+1}\frac{3^j}{j})=\frac{2}{3^j}(j=1,2,3,\cdots),$$

由于级数

$$\sum_{j=1}^{+\infty}|(-1)^{j+1}\frac{3^j}{j}\cdot\frac{2}{3^j}|=2\sum_{j=1}^{+\infty}\frac{1}{j}$$

发散，故随机变量X的数学期望不存在，这是构造概率中反例时的一个例证。

二项分布$B(n,p)$的数学期望与方差的计算，可以借助于二项分布展开式，运算中显示了一定程度的复杂度：

设随机变量$X\sim B(n,p)$，则数学期望

$$E(X)=\sum_{k=0}^{n}kC_n^k p^k q^{n-k}$$
$$=\sum_{k=1}^{n}\frac{n(n-1)\cdots(n-k+1)}{(k-1)!}p^k q^{n-k}$$

为了向二项式靠拢，从求和号中提出因子 p，使其变动着的下标“向 $k-1$ 看齐”，则有

$$E(X)=p\sum_{k=1}^{n}\frac{n(n-1)\cdots(n-k+1)}{(k-1)!}p^{k-1}q^{(n-1)-(k-1)},$$

原来求和号中的 q^{n-k}，在“向 $k-1$ 看齐”后，变成了 $q^{(n-1)-(k-1)}$，显见，下一步的化简要领应是“向 $n-1$ 看齐”，为此，需从求和号中提出因子 n，并适当调整求和号的上下标，可得

$$E(X)=np\sum_{k=0}^{n-1}\frac{(n-1)\cdots[(n-1)-(k-1)]}{k!}p^k q^{(n-1)-k}$$
$$=np\sum_{k=0}^{n-1}\frac{(n-1)\cdots[(n-1)-(k-1)]}{k!}p^k q^{(n-1)-k}$$
$$=np(p+q)^{n-1}np.$$

同理

$$E(X^2)=E[X(X-1)]+E(X)$$
$$=\sum_{k=2}^{n}k(k-1)C_n^k p^k q^{n-k}+np$$
$$=\sum_{k=2}^{n}\frac{n(n-1)\cdots(n-k+1)}{(k-2)!}p^k q^{n-k}+np$$
$$(\text{向}k-2\text{看齐})=p^2\sum_{k=2}^{n}\frac{n(n-1)\cdots(n-k+1)}{(k-2)!}p^{k-2}q^{(n-2)-(k-2)}+np$$
$$=n(n-1)p^2\sum_{k=2}^{n}\frac{(n-2)(n-3)\cdots[(n-2)-(k-3)]}{(k-2)!}p^{k-2}q^{(n-2)-(k-2)}+np,$$

调整$\sum$的上下标 $=n(n-1)p^2\sum_{k=0}^{n-2}\frac{(n-2)(n-3)\cdots[(n-2)-(k-1)]}{k!}p^k q^{(n-2)-k}+np$

（二项式定理）
$$=n(n-1)p^2(p+q)^{n-2}+np$$
$$=n(n-1)p^2+np,$$

从而

$$D(X)=E(X^2)-E^2(X)$$
$$=n(n-1)p^2+np-(np)^2=npq.$$

在计算几何分布的期望与方差时，需运用密级数的性质，交换求和号与求导符号的次序，这代表着概率计算中的一种特殊技巧。

设随机变量 X 服从几何分布，其分布率为

$$P(X=K)=pq^{k-1}(k=1,2,\cdots),$$

则数学期望

$$E(X)=\sum_{k=1}^{+\infty}kpq^{k-1}=p\sum_{k=1}^{+\infty}kq^{k-1}=p(\sum_{k=1}^{+\infty}q^k)'=p(\frac{q}{1-q})'$$

$$=p\frac{1}{(1-q)^2}=\frac{1}{p},$$

$$E(X^2)=\sum_{k=1}^{+\infty}k^2pq^{k-1}=p\sum_{k=1}^{+\infty}k^2q^{k-1}=p(\sum_{k=1}^{+\infty}kq^k)'=p(q\sum_{k=1}^{+\infty}kq^{k-1})'$$

$$=p[\frac{q}{(1-q)^2}]'=p\frac{1+q}{(1-q)^3}=\frac{1+q}{p^2},$$

所以

$$D(X)=E(X^2)-E^2(X)=\frac{1+q}{p^2}-\frac{1}{p^2}=\frac{q}{p^2}. \quad (7\text{–}2)$$

著名数学家高斯逝世于1856年，高斯在格丁根大学的教授职位，由狄利克雷继任；1859年，狄利克雷去世，接替狄利克雷的教授席位的，是年仅33岁的黎曼。黎曼也是高斯的得意门生，同时是狄利克雷、雅可比等大数学家的学友。德国数学家雅可比是柏林大学的哲学博士，他在数学分析领域的杰出贡献，较为集中地体现在他引进的“雅可比行列式”上，应用雅可比行列式J，可以一揽子解决多维随机变量(X,Y)的函数$Z(X,Y)$的概率分布问题。

不妨以简单的情形说明之。设二维随机变量（X，Y）的概率密度函数为$f(x,y)$，

$$Z=Z(X,Y)=X+Y,$$

则有

$$F_Z(z)=\iint_{D:X+Y\le z}f(x,y)dxdy, \quad (7\text{–}3)$$

仔细考虑区域D的边界，将二重积分化为二次积分后得

$$F_Z(z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x}f(x,y)dydx$$

$$=\int_{-\infty}^{z}\int_{-\infty}^{+\infty}f(x,u-x)dxdu,$$

所以可知

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx.$$

若引入雅可比行列式，进行积分的变量替换，则可以令

$$\begin{cases}u=x+y\\x=x\end{cases},$$

经过计算得

$$J=\frac{U(x,y)}{U(x,u)}=\begin{vmatrix}1&0\\-1&1\end{vmatrix}=1,$$

于此可得与前述完全相同的$F_Z(z)$与$F_Z(z)$的结论。

这似乎表明，在函数$Z=Z(X,Y)$较为简单的情况下，用不用雅可比行列式进行变量替换，难易程度是差不多的。当$Z=Z(X,Y)$的表达式稍微复杂一些，雅可比行列式的作用将很显著。

现在假设$Z=Z(X,Y)=XY$，

此时令

$$\begin{cases}u=xy\\x=x\end{cases},$$

即

$$\begin{cases}x=x\\y=\dfrac{u}{x}\end{cases},$$

则有

$$J=\frac{U(x,y)}{U(x,u)}=\begin{vmatrix}1 & 0\\-\dfrac{u}{x^2} & \dfrac{1}{x}\end{vmatrix}=\frac{1}{x},$$

进而得

$$F_Z(z)=\int_{-\infty}^{z}\int_{-\infty}^{+\infty}f(x,\frac{u}{x})|\frac{1}{x}|dxdu,$$

所以可知

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,\frac{z}{x})\frac{dx}{|x|}。$$

若假设 $Z=Z(X,Y)=X/Y$，此时令$\begin{cases}u=x/y\\y=y\end{cases}$，即$\begin{cases}x=uy\\y=y\end{cases}$，
则有

$$J=\frac{U(x,y)}{U(u,y)}=\begin{vmatrix}y & u\\0 & 1\end{vmatrix}=y,$$

进而得

$$F_Z(z)=\int_{-\infty}^{z}\int_{-\infty}^{+\infty}f(yu,y)|y|dydu,$$

所以可知

$$f_Z(z)=\int_{-\infty}^{+\infty}f(yu,y)|y|dy \tag{7-4}$$

由于不必绘出区域 D 的图形，因而积分限的颠倒错乱的顾虑几乎可以打消。具体运算时可以无视直观背景，直接对照雅可比行列式作代数推演，这就显示了雅可比行列式的重要性和简捷性，这种简捷性将随着函数形式的复杂化，显示得愈加清晰。

极限论构成了数学分析的基础，微积分中一系列重要的概念和方法，都与极限关系密切。概率论中运用极限的地方非常多，现以分布函数的性质与极限定理为例说明之。

对分布函数 $F(x)$ 来说，有如下两个极限：

（1）$\lim\limits_{x\to+\infty}F(x)=\lim\limits_{x\to+\infty}P(X\le x)=1$　。

（2）$\lim\limits_{x\to-\infty}F(x)=\lim\limits_{x\to-\infty}P(X\le x)=0$　。

如果仅凭直觉，这两个极限是显然的，因为极限值分别是必然事件 Ω 与不可能事件 Φ 的概率，当然是 1 与 0。但若是严格证明，则需借助数学分析中的许多东西。

由概率的可列可加性

$$P(-\infty<X<+\infty)=\sum_{n=-\infty}^{+\infty}P(n<X\le n+1)=\lim_{N\to+\infty}\sum_{n=-N}^{N}[F(n+1)-F(n)]$$

$$=\lim_{N\to+\infty}F(N)-\lim_{N\to-\infty}F(N)=\lim_{x\to+\infty}F(x)-\lim_{x\to-\infty}F(x).$$

因为

$$P(-\infty<X<+\infty)=P(\Omega)=1,$$

故有

$$\lim_{x\to+\infty}F(x)-\lim_{x\to-\infty}F(x)=1.$$

又因为$F(x)$单调上升且存在上、下界，即

$$0\le F(x)\le 1.$$

由微积分的知识可知：

(1 $\lim_{x\to+\infty}F(x)$　$\lim_{x\to-\infty}F(x)$　均存在。

(2 $F(x)$　能够达到其上确界 1、下确界 0。

由极限存在的唯一性，综合起来得

$$\lim_{x\to+\infty}F(x)=1,\ \lim_{x\to-\infty}F(x)=0.$$

综上所述，数学分析的思想方法已经渗透到了概率论的各个方面。没有微积分的推动，就没有概率论的公理化与系统化，概率论就难以形成一门独立的学科。数学分析与概率论的亲缘关系，决定了概率论的确定论的特征。由此可见，概率论具有线性性与非线性性的双重特点，是一门同时包含着确定性和非确定性二重品格的特殊的数学学科。

7.6 概率论在统计学发展中的角色研究

“统计”现在已经是一个很流行的名词，在当今信息时代，各国政府经常发布有关国情的数据资料，如某一时期经济增长率和失业率等。一些非官方机构，如工会、商会、大学及专门学会、新闻机构也在特定的领域内从事收集整理和发布数据的工作。统计工作只是在近代随着种种条件的改进才变得日趋完善，其比较原始的形态却有着非常悠久的历史。我国古籍中常见有关人口、口粮以及地震和水旱灾等记录。在西方，据记载，在共和罗马时期，4 年一度对每个家庭的人口和财产进行普查登记，当然，这种活动与统计学作为一门学科的建立还不能画等号，但其促进作用是无可否认的。学者们指出，现今流行的“统计学”（statistics）一词源自意大利文 stato，其词根兼有"国家"和"情况"的意义。“统计学家”（statistician）一词源自意大利文 statista，当时理解为“处理国务的人”，因而当时的统计学就是国情学，这种思想流行于 16 世纪的意大利，后来传播到其他欧洲国家。

大量的原始数据如果不经过整理、分类、排比、分析，并通过适当的形式表示出来，就好比一堆没有经过冶炼的矿物，没有任何价值。对数据的整理分析工作很早就有人开始研究了，系统地从事这一工作并对后世统计学发展有重大影响的是英国学者格朗特，他在 1662 年发表的《关于死亡公报的自然和政治观察》一书是关于统计的开山之作，该著作有以下几个主要的创新思想：

（1）提出了“数据简约”的思想，即把大量的、杂乱无章的数据，依某种分类标准，整理成一些意义明晰的表格，使数据中包含的有用信息凸现出来。这样一种思想，直到现今仍被统计学家视为基础性的工作。

（2）指出了数据的可信性问题，即数据是否真实，是否有人出于某种目的而对数据进行了篡改，或在数据的获取过程中出现了重大失误，如仪器未调准或登录时书写有误等。数据中这样的数值称为异常值，鉴别数据中是否有及何者为异常值，直到今天仍是一个在应用上很重要并在方法研究上受到重视的问题。

（3）统计比率的稳定性概念，指某种特性出现的频率随着观察次数的增加而趋于稳定。格朗特在书中并未用明确的语言给出这一原则，而是通过对具体数据的处理显示了他的统计分析是基于这样一个原则。

（4）引进了生命表的概念，生命表是指现存人口的年龄分布，利用生命表可计算在某一年龄间隔内的人数的百分比，可计算一个活到某一年龄 t 的人至少再活 Δt 年的百分比，而这对于保险金、年金等的计算有直接的关系，这个概念对后世人口统计学有很大的影响。

格朗特的人口统计工作在欧洲很有影响，促成一些国家建立了政府统计部门，特别值得一提的是，他的工作影响了佩蒂，促使他建立其"政治算术"，佩蒂是英国 17 世纪的政治经济学家，1676 年写成了其政治算术的代表作《政治算术》。佩蒂主张用统计数字来分析政治、经济和社会问题，国家的一项政策效果如何，不能靠口舌辩论，也不能靠抽象的、看似灵巧的推理，一切让数字说话，他的思想受到英国伟大的科学与哲学家培根的很大影响，培根的实证科学思想，主张科学理论应以实际观察为依据并接受其检验。佩蒂对统计方法的贡献甚为有限，他对统计学的最大贡献在于开拓了统计方法的应用面，使统计研究不再局限于人口统计问题。

奎特勒是 19 世纪最有影响的统计学家之一，他的主要贡献是倡导并身体力行将正态分布用于连续性数据的分析，他的这一努力使正态分布律推广到其他数据，并发明了一种方法，将一批数据拟合于正态分布。他用这个方法处理过许多数据，多数都拟合较好，他倡导用统计方法来定量研究社会规律，提出了"社会物理学"的大胆设想，即随着对支配社会的规律的量化研究，这些规律有朝一日可以达到像天文学和物理学那样精密的程度。这种设想严格意义上是难以实现的，不过现代社会科学的数量化趋势确实在与日俱长，有些社会科学分支，特别是经济学、人口学、环境学都用到很深的数理统计理论，这也不妨可看做是奎特勒设想的部分实现。

现代意义上使用的“统计学”一词，是英国学者辛克莱在 18 世纪末提出的。随后，许多应用统计学家和理论学家对统计方法及理论进行了深入广泛的研究：凯尔在 19 世纪末提出了“代表性抽样”的思想，即样本必须具有代表性，在此前提下，并不需要特别大的样本量，就可以得到总体指标的较满意估计；鲍莱把概率方法引进到抽样调查中，采用随机抽样方法，在理论上验证了抽样方法的合理性，并撰写了专著《抽样调查精度的度量》；费歇尔自 1919 年起在英国一个农业实验站工作了十多年，从事农业试验及其统计分析的研究工作，提出在随机化的设计中融入系统性因子的作用以降低随机化带来的误差；著名英国统计学家皮尔逊、印度统计学家马哈拉诺比斯对统计学都作出过重要贡献。自 20 世纪 30 年代以来，抽样调查方法受到包括美国在内的一些国家的重视，使统计学的应用变得十分广泛。

古典统计学主要有两大分支：国势学与政治算术。就国势学而言，它产生于时代的需要，由于资本主义发展的不平衡，各国国情不同，统治阶级治理国家需要具备国内外的知

识，其中包括一些重要的统计数据。因此，它是为了满足政治活动家的需要。而它研究的方法主要是文字记述，更不用说探索事物之间的关系。

作为观点统计学的真正源头——政治算术，通过统计资料的分析研究以发现社会现象中的统计规律性，为国家制定政策提供可靠的统计资料。其研究方法也与国势学有很大区别，采取了计量方法来研究社会经济现象，同时采用了图表法来分析和反映统计资料，与图表法紧密结合在一起的有分组法，已体现了按品质标准分组和按数量标志分组。除了前三种方法，还有推算方法，在当时统计资料不多的情况下，遵循事物固有的依存关系，工具少量资料教学推算，这在间接计算史上具有重大意义。政治算术将各各种学科，如数学、逻辑学、经济学、会计学等引入统计学，同时将先进的科学方法，如大量观察法、图表法、分组法、比较法、比例法引入统计领域，构成古典统计学的方法体系，奠定了真正的统计学基础。

古典统计时期，概率论基本上是独立发展的。换言之，它与统计学的联系不多。但这个时期的学者们已有将概率论应用于社会现象研究的种种设想经过一段时间的发展，随着概率论的不断发展，逐步步入社会科学领域，且与统计理论的不断结合，为进一步研究现象的内在特征打下坚实的基础。统计已较明显地体现了“收集和分析的量”的特征，统计实证方法和理论方法浑然一体，它们的有机结合显示出统计的广阔前景。

近代他继续是一个由古典统计学向现代统计学过渡的阶段，它是古典统计学的继续和发展。其中国势学派继续分化，故步自封者日趋没落，而吸取新营养者则朝着新的方向发展，也就是与政治算术合流，向现代统计学发展。

对于国势学派而言，它仍然局限于“国家显著事项”的范围之中，仍然以文字记述为主。它虽然有一定程度的发展，但还是慢慢地趋于没落。当时在国势学和政治算术之间形成了政府统计，兼顾了二者的一些特点，有一定的代表性。而政治算术学派则有了一定程度的发展。由于经济的发展，有些学者把统计应用于人口调查、社会调查、保险业、公共卫生事业等等，形成了经济统计、农业统计、物价指数等。统计在社会发展所取作用越来越明显，为当时社会许多学者所认同。但政治算术无论是图表还是比较派都是以描述为主，并没有历史性的突破，其结果不过是历史的再现。

虽然古典统计时期的概率论基本上独立发展的，它与统计的联系不多。但进入近代统计时期，由于在古典统计时期有将概率论用于研究社会现象的种种设想，越来越多的人开始用概率论研究法律、政治、道德、经济等社会问题，并且有许多数学家和统计学家为统计的发展作出了重大贡献，其中典型代表人物有凯特勒和拉普拉斯。突出的学术成果有几何概率论、伯努利定理、最小二乘法、拉普拉斯变换、误差理论、大数法则和正态分布等等。早期概率被人们认为，它既可以看做是某个内在特性，这个特性不依赖于我们对该系统的知识；也可以看做是某一陈述相信程度的度量。概率论在这个时期取得长足的发展。

概率论的数学定律是客观存在于宇宙中的大量随即现象的统计规律性的反映。它应用数学方法来研究这种随即现象，且由于它的特有的方法而使它成为数学的一个分支。正如凯特勒所说“要更多地促进科学方发展，必须使之更多地进入数学领域，这是一种集中趋势。我们判断一门科学的精密程度，就看它应用数学的多寡”。当然统计学也不例外。正因为概率论引入统计学，才使概率论中最基本的概念随机事件和概率引入统计之中，这样在处理大量的统计资料时带来方便，以概率论为基础创造一套如何对客观现象进行搜集资

料，整理分析以及推断的方法，做到了根据样本推断总体，从具体到一般。拉普拉斯以大数法则为桥梁，将概率论与政治算术联系起来。在 19 世纪中叶，数理统计学派的奠基人比利时物理学家和统计学家凯特勒，将自然科学的研究精神和方法应用到社会领域。他最先用大数定律论证了社会生活现象纷繁复杂变化的偶然性中存在着规律性，并且提出了误差理论，用来解决统计学上准确性问题。概率论在这个阶段为统计学的发展铺设了一个坚实的基础，为现代统计学的进一步发展做好准备。

随着概率论的进一步发展，在引进概率论的基础上统计学有了质的飞跃，表现为由大样本理论向小样本理论发展，即由描述统计学向推断统计学发展。

任何科学都是揭示事物特殊矛盾运动规律。现代统计经过一段时间发展成为一门较为成熟的研究关于数字资料收集、组织、分析和解释的科学。其中资料收集是取得数量或数据的方法。正确的统计结论来源于正确的资料，以得出符合逻辑的结论，而表格或图形是用来表现和组织资料的典型方法。资料分析是从给定的量中抽出有关问题，从而得出一个简单的综合性的结果，而达到这个目的重要的有平均数、中位数、极差、标准差等。资料解释是通过资料分析来做出结论的工作，它通常是通过类似现象小的集合提供的信息来对有关对象大的集合形成预测。

概率论是在随即现象的一般数学模型的基础上研究随机事件、概率和随机变量等的基本规律，找出其内在的性质和联系。其研究的主要内容是：概率的运算、大数法则、极限分布和随机过程。而统计则是以概率论为一个支点，创造一套如何对客观现象资料进行搜集、整理、分析及推断的方法。概率论用的是演绎法，从一般到具体。如它通过对一些典型概率模型的研究，应用于客观实际，从而对一些具体的客观随即现象可能出现的结果做出估计。而统计用的是归纳法，它通过对客观现象部分资料的观察、搜集、整理和分析，根据样本推断总体，从具体到一般。因此，统计学应以概率论为理论基础，而概率论则需要统计在实践中加以验证。

如果古典统计学的功能主要是描述事物的局部和现状的话，那么现代统计学的功能主要是推断事物的总体和未来。因此，现代统计学比较古典统计学是一个巨大的进步，或者说是一个质的飞跃。作为飞跃的主要标志之一是统计推断，推断过程大都是依据包含在样本数据中的不完备信息，它必然要冒风险，原因在于不确定性，而概率就是这种不确定性的计量。因此概率论是统计推断的指导思想和理论基础。可以说没有概率论就没有统计推断，没有统计推断就没有现代统计。

（1）概率论促进统计的定量研究。

概率论提高了统计数据的代表性。统计工作的开展往往需要大量的统计数据，而这些统计数据来自于各种统计工作方法，其结果是否正确不但取决于观察数据的多寡，而且取决于取得正确代表性的方法。例如，建立在随机原则基础上的抽样调查，其所耗费用低，取得资料速度快，而且应用广泛。只有经过精心设计和组织，抽样调查效果可能超过全面调查水平，这样就大大提高了统计数据质量的准确性和统计工作的效率，同时拓展了统计研究对象的数量信息内涵。

概率论加强了统计数据的抽象描述。在统计学中矩、多元统计分析、时间序列分析等在统计数据的定量处理占重要地位。矩本来是物理学中的一个名词，但均数和方差同重心和转动惯量之间的类似使得使用矩来描述均数和方差。作为以均值为基础而定义的数字特

征，矩把统计数据的两个基本数量特征（平均数和方程）综合到一个概念，它体现了统计求同察异的思想，即在了解差异的同时认识事物的同质性。在 19 世纪末叶俄国数学家切比雪夫创立一般性的大数定律，即：DE 抽样数量愈大，则抽样总体的平均数与全部平均数之间的离差愈小，而概率接近 1，并且拟定了“动差法”，即：估计随机变量的平均值和方差的方法。这可以看作矩思想的早期体现。多元统计分析是针对横截面数据而作的相关和回归分析，研究多个随机变量之间相互依赖关系以及内在统计规律性，提高了统计定量分析的科学性和可靠性。时间序列分析研究纵剖面数据所表现的现象以及现象之间关系的发展变化规律，从纵向显示其描述统计数据的有效性和重要性。多元统计分析和时间序列分析不仅在统计描述方面发挥重要作用，而且在统计推断方面也发挥了重要的作用。

（2）概率论使统计由描述阶段向推断阶段发展。

在古典统计学阶段和近代统计学阶段，统计理论和概率论在广泛应用中有了很大的发展，其基本上是由描述统计阶段向推动统计阶段发展，其中概率论起着重要的作用。这一发展阶段其反映了统计学发展的巨大成就，也是统计学成熟的重要标志。统计所研究的统计规律性要通过大量观察才被发现。但是在客观上有时只允许我们对随机现象进行有限次观察，因此推断总体就会产生误差。这样我们需要对误差做出估计，否则这种推断也就没有意义。运用概率论的原理，根据整体和部分之间的内在联系进行分析和推断，就可以使归纳推断产生的不确定性得到度量和控制。1763 年，英国贝叶斯发表的贝叶斯定理，给出来已知结果 E，对所有原因 C 计算其条件概率（后验概率）$P_E(C)$ 的公式，这可以看作最早的一种统计推断程序。拉普拉斯和高斯等利用贝叶斯公式估计参数，高斯在建立了以“最小二乘法”为基础的误差分析，这一系列科研成果使统计学摆脱了对观察数据的单纯描述而向推断过渡。

推断统计实质是在随机抽样的基础上，通过样本特征数来认识总体特征数，其主要问题是参数估计和假设检验。对于参数估计而言，由于样本的同质性，以样本的数据在一定概率条件下来估计参数，它是依据样本数据对总体参数作猜想，其实质上是一种类比。点估计借助于总体的一个样本来估计总体分布中的未知参数的值，其衡量标准主要有无偏性、有效性和一致性。但有时就点估计再估计出一个范围，并希望知道包含未知参数真值的可信程度，这样的范围通常以区间表示，同时还给出了此区间包含未知参数真值的可信程度，这就是区间估计。而假设检验的基本思想是应用小概率原理，小概率原理指发生概率很小的随机事件在一次试验中是几乎不可能发生的。根据这一原理对我们所关心的却又是未知的参数先作假设，然后抽取样本，利用样本提供的信息对的正确性进行判断的过程。无论参数估计还是假设检验都是以概率论为基础，都展现出现代统计质的飞跃。

总之，随着社会的发展，概率论逐渐进入统计领域，使得其内涵的研究抓住了事物的随机性和确定性，透过随机性去研究事物的发展规律。同时统计学在应用中其研究的对象也日益扩大，它既继承传统领域的研究，又拓宽了研究范围，对社会经济、自然现象和心理思想进行研究，形成了一门应用范围广泛的综合性方法论科学。

7.7 最小一乘法与最小二乘法——历史与差异

1.引言

设有 n 个数 $a_1,a_2,\cdots,a_n$，要找一个数 x 反映这组数的总的情况，使得 x 和这 n 个数的

偏差

$$x-a_1,x-a_2,\cdots,x-a_n \tag{7-5}$$

在总体上说来尽可能地小。

采用的常见方法就是最小二乘法准则，即选取适当 x 值，使得这些偏差的平方之和

$$D=\sum_{i=1}^{n}(x-a_i)^2 \tag{7-6}$$

达到它的最小值。

令
$$\frac{dD}{dx}=2\sum_{i=1}^{n}(x-a_i)=0,$$

得
$$\bar{x}=\frac{a_1+a_2+\cdots+a_n}{n}$$

此 $\bar{x}$ 即为 n 个数 $a_1,a_2,\cdots,a_n$ 的算术平均数。

为什么不把式 7-5 的 n 个偏差本身相加，而把它们的平方和相加，这是因为这些偏差有些为正值，有些为负值，直接相加，就会正负相消，不能反映总体的情况。

有人问能否把式 7-5 的 n 个偏差的绝对值相加，即找一个数使得

$$D=\sum_{i=1}^{n}|x-a_i| \tag{7-7}$$

最小?

回答是肯定的，这就是最小一乘法准则。所求的这个数就是中位数 med：

（1）当 n 为偶数时，中位数 $med[a_{(\frac{n}{2})},a_{(\frac{n}{2}+1)}]$ 中的任何一个数。

（2）当 n 为奇数时，中位数 med 是 $a_{(\frac{n+1}{2})}$ 这个数。这里 $a_{(1)}\le a_{(2)}\le\cdots\le a_{(n)}$ 是 $a_1,a_2,\cdots,a_n$ 按由小到大的排列（统计学上称 $a_{(1)},a_{(2)},\cdots,a_{(n)}$ 为 $a_1,a_2\ \cdots,a_n$ 的顺序统计量）。

由上面的分析可知，算术平均数 $\bar{x}$ 是到 n 个数 $a_1,a_2,\cdots,a_n$ 的距离的平方和最小，而中位数 med 是到 n 个数 $a_1,a_2,\cdots,a_n$ 的距离的和最小，这才是算术平均数和中位数的实质。

对于二维情形，已知两点 (x_1,y_1)，(x_2,y_2) 可确定一条直线，$y=a+bx$ 这只需将两点坐标代入直线方程，解出 a,b 即可。将两点推广到 n 个点，如何确定线性回归直线呢?

最常用的是利用最小二乘法准则，即求出 a,b，使得 (x_1,y_1)，(x_2,y_2) ，$\cdots$，(x_n,y_n) 各点沿 y 轴到直线 $y=a+bx$ 的偏差的平方之和最小，即

$$f(\beta)=\sum_{i=1}^{n}[y_i-(a+bx_i)]^2=\min \tag{7-8}$$

对上式中的 a,b 分别求偏导并令其为零，若其系数行列式不为零，则可求得唯一的 a,b 的显式表达，它们即为所求的回归系数

$$\begin{cases}\hat{a}=\bar{y}-\hat{b}\bar{x}\\ \hat{b}=\dfrac{l_{xy}}{l_{xx}}\end{cases}, \tag{7-9}$$

其中$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ ，$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ ，$l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2$ ，$l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$

那么能否给出另一种准则，求出 a,b，使得(x_1,y_1) ，(x_2,y_2) ，…，(x_n,y_n)各点沿 y 轴到直线$y=a+bx$ 的偏差的绝对值之和最小，即

$$f(\beta)=\sum_{i=1}^{n}|y_i-(a+bx_i)|=\min ? \qquad (7\text{-}10)$$

答案是可行的，此即为最小一乘法准则。

也许是计算上的困难，一般只用最小二乘法，很少涉足最小一乘法。1805 年，法国数学家勒让德在研究天文学和测地学处理数据时最先发明最小二乘法，后来高斯等数学家对最小二乘法进行了大量的理论研究和应用，在统计学中发挥着重要的作用，是十九世纪统计学的“中心主题”。但从历史上看最小一乘法比最小二乘法的起源要早 40 多年。自 1755 年起，波斯科维奇投身于子午线长的问题研究，在 1760 年提出来最小一乘法准则。但由于其计算上的困难及其它原因，对这个准则的研究长期处于停止状态，直到 20 世纪 50 年代以后，局面才改变，越来越受到统计学家的重视。

2.最小一乘法的计算

首先对 a,b 加以约束，使得回归直线$y=a+bx$ 经一给定的点(x_1,y_1) ，即

$$y_1=a+bx_1 ,$$

在此条件下，使 7-10 式达到最小。

对给定的点(x_1,y_1) ，作变换

$$\begin{cases} x_i'=x_i-x_1 \\ y_i'=y_i-y_1 \end{cases} (i=1,2,\cdots,n), \qquad (7\text{-}11)$$

这样式 7-10 就转化为求 b，
使得

$$f(b)=\sum_{i=2}^{n}|y_i'-bx_i'|=\min ,$$

$f(b)$ 是关于 b 的线性折线凸函数，在 b 轴上，每个折点坐标为 y_i'/x_i' $(i=2,\cdots,n)$（其中 y_i'/x_i' 是$|y_i'-bx_i'|$的最小值点）。

不失一般性，可设 $x_i'\neq 0$ $(i\geq 2)$。事实上，$x_i'=0$，则$|y_i'-bx_i'|=|y_i'|$与 b 无关，因此，这一项在确定 b 的过程中不起作用，故可将其剔除，其余 x_i' 均不等于 0。按 y_i'/x_i' 的大小排序

$$y_{i_2}'/x_{i_2}'\leq\cdots\leq y_{i_n}'/x_{i_n}' ,$$

由此得到 x_i' 的一个排序

$$x_{i_2}', x_{i_3}',\cdots, x_{i_n}' .$$

记

$$b_{(k)}=y_{i_k}'/x_{i_k}' \ (k=2,\cdots,n),\ M=\sum_{k=2}^{n}|x_{i_k}'|,$$

若

$$\sum_{k=2}^{r}|x'_{i_k}|\ge M/2 \text{且} \sum_{k=2}^{r-1}|x'_{i_k}|<M/2,$$

则回归直线为

$$y-y_1=\frac{y'_{i_r}}{x'_{i_r}}(x-x_1).$$

记

$$b_1=y'_{i_r}/x'_{i_r},$$

则 $f(b)$ 值（即偏差绝对值之和）为 $f(b_1)$。

一般说来，上面所求的回归直线，不妨设为 $y=a_1+b_1x$ ，会通过所给数据中的另一点，设为 (x_2,y_2) 作与式 7-11 类似的变换，重复上述步骤，可求得最优的 a_2,b_2 和对应值 $f(b)=f(b_2)$。因为 $y=a_1+b_1x$ 也是通过 (x_2,y_2) 的战线，所以 $f(b_2)$ 必定小于 $f(b_1)$。如此重复以上步骤，直到找到一点 (x_t,y_t)，使得经过它的回归直线也通过 (x_{t-1},y_{t-1})，对应值 $f(b)=f(b_t)$ 最小。这条直线所表达的方程就是最终要求的最小一乘法回归直线。

3.实证分析

对表 7~1 中的数据用最小一乘法进行回归分析。

表 7-1

i	1	2	3	4	5	6	7	合计
x_i	1	2	3	4	5	6	7	28
y_i	49	95	146	220	261	303	438	1512

取 $(x_1,y_1)=(1,49)$，

因为

$$M=21,\sum_{k=2}^{4}|x'_{i_k}|=8<M/2=10.5,\sum_{k=2}^{5}|x'_{i_k}|=12>M/2=10.5,$$

所以

$$b=b_{(5)}=53.0,$$

回归直线为

$$y-49=53(x-1),$$

即

$$y=53x-4,$$

该直线过（5，261）（表 7-2）。

表 7-2

x_i	1	2	3	4	5	6	7	合计
y_i	49	95	146	220	261	303	438	1512
$x'_i=x_i-1$	0	1	2	3	4	5	6	21

续表 7-2

$y_i'=y_i-49$	0	46	97	171	212	254	389	1169
y_i'/x_i'		46.0	48.5	57.0	53.0	50.8	64.8	
$b_{(k)}$		46.0	48.5	50.8	53.0	57.0	64.8	
$\lvert x_{i_k}'\rvert$		1	2	5	4	3	6	21

取$(x_2,y_2)=(5，261)$，

由于

$$M=13，\sum_{k=2}^{4}|x_{i_k}'|=6<M/2=6.5，\sum_{k=2}^{5}|x_{i_k}'|=9>M/2=6.5，$$

故

$$b_1=b_{(5)}=55.3，$$

回归直线为

$$y-261=55.3(x-5)，$$

即

$$y=55.3x-15.6，$$

该直线过（2，95）（表 7-3）。

表 7-3

x_i	5	1	2	3	4	6	7	合计
y_i	261	49	95	146	220	303	438	1512
$x_i'=x_i-5$	0	–4	–3	–2	–1	1	2	–7
$y_i'=y_i-261$	0	–212	–166	–115	–41	42	177	–315
y_i'/x_i'		53.0	55.3	57.5	41.0	42.0	88.5	
$b_{(k)}$		41.0	42.0	53.0	55.3	57.5	88.5	
$\lvert x_{i_k}'\rvert$		1	1	4	3	2	2	13

取 $(x_3,y_3)=(2，95)$，

由于

$$M=16，\sum_{k=2}^{4}|x_{i_k}'|=6<M/2=8，\sum_{k=2}^{5}|x_{i_k}'|=9>M/2=8，$$

故

$$b=b_{(5)}=55.3，$$

回归直线为

$$y-95=55.3(x-2)，$$

即

$$y=55.3x-15.6，$$

该直线过（5，261）（表 7-4）.

表 7-4

x_i	2	1	3	4	5	6	7	合计
y_i	95	49	146	220	261	303	438	1512
$x_i'=x_i-2$	0	-1	1	2	3	4	5	14
$y_i'=y_i-95$	0	-46	51.0	125.0	166.0	208.0	343.0	847
y_i'/x_i'		46.0	51.0	62.5	55.3	52.0	68.6	
$b_{(k)}$		46.0	51.0	52.0	55.3	62.5	68.6	
$\lvert x_{i_k}'\rvert$		1	1	4	3	2	5	16

因此，根据表 7-1 中的数据，经过 3 次迭代，用最小一乘法准则所求的回归直线为，$y=55.3x-15.6$ 其偏差绝对值之和为

$$f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx$$

而根据表 7-1 中的数据，用最小二乘法准则式 7-9 可得回归直线为

$$y=60.6x-26.6,$$

其偏差平方之和为

$$f(\beta)=\sum_{i=1}^{n}[y_i-(a+bx_i)]^2=118.6>107.7\cdot$$

可见最小一乘法回归模型优于最小二乘法模型。与最小二乘法比较，最小一乘法具有以下特点：

（1）异常值对回归线的影响：在回归分析的最小二乘法估计中，远离中心的突出点，对参数的计算结果有较大的影响。而在最小一乘法估计中，异常值的影响一般较小。

（2）解的多重性：回归分析中用最小一乘法，有时会出现多个解，没有显式表达。而对于最小二乘法，解一般是唯一的，且有显式表达。

不过，有人做过这样的实验：拿大量的 $x-y$ 散点图，让一些人各自用目测的方法配直线。结果表明，大多数人目测的结果更接近于最小一乘法而不是最小二乘法。

第 8 章　概率统计教学中渗透概率统计发展史

8.1 哲学与文化视角下概率统计课的育人功能

概率论与数理统计是认识和理解随机世界的一把钥匙。从哲学、文化视角来看，概率论与数理统计具有无价的文化内涵和育人功能，它是进行唯物辩证法教育的好素材；具有辩证唯物主义认识论教育功能；在培养人的科学品质方面具有得天独厚的功能。在概率论与数理统计教学中注重挖掘这些因素，对实施素质教育有着十分重要的意义。著名数学家拉普拉斯说："生活中最重要的问题，其中绝大多数在实质上只是概率问题"。概率论是"生活真正的领路人，如果没有概率的某种估计，我们将寸步难移，无所作为"。（W.S.Jevons）

信息化是当今时代一大潮流，而"信息"基本上是以数据的形式来反映的。数理统计是专门处理那种受到随机因素影响的数据的有力武器。一句话，概率论与数理统计是认识和理解随机世界的一把钥匙。对随机性的认识，是每一个信息时代的公民的知识结构中应具备的成分，也是一个人的人文素质的一部分，学习概率论与数理统计在信息社会里越来越重要。然而，传统教学往往重视的是概率论与数理统计的科学工具性价值，看到的只是它在技术层面的功能，而忽略了概率论与数理统计中无价的文化内涵与文化层面的育人功能。为了有效实施素质教育，我们从哲学、文化视角窥视概率论与数理统计课的人文素质教育因素与功能，引起教学的重视，这对于深化这门课的教学改革，加大概率论与数理统计教学在素质教育中的贡献力度具有十分重要的意义。

1.进行唯物辩证法教育的好素材

"从哲学意义上讲：人的素质最为核心的是他的世界观和方法论"，概率论与数理统计的知识内容中蕴含着丰富的辩证哲理，是进行唯物史观与辩证法教育的好素材，可以帮助学生树立和形成辩证唯物主义的世界观和方法论。

(1)偶然性与必然性是哲学上的一对矛盾范畴，它们相互依存、相互制约、相互转化，必然性寓于偶然性之中。恩格斯指出："在表面偶然性起作用的地方，这种偶然性始终是受内部隐蔽的规律支配的。而我们的问题只是在于发现这些规律"。

随机现象有其偶然性一面，也有其必然性一面，这种必然性表现在大量重复试验或观察中呈现出的固有规律，称之为随机现象的统计规律，而概率论是演绎推理随机现象的统计规律；数理统计则是以概率的理论为基础，通过实验、归纳、推理随机现象的统计规律。概率论与数理统计的研究目的正是在于从偶然性中探求必然性，从混沌中寻找有序，该学科本身体现了偶然性与必然性的辩证关系。

(2)频率与概率是两个对立的概念，事件的概率是固定值（常数），事件的相对频率则是一个与试验次数有关的一组来回振荡的数（变数）。而我们把概率看做是频率的稳定值（即概率意义上的极限值，并非通常数学中的极限值），这是因为，无论试验次数多么大，都无法保证事件的频率值充分地接近事件的概率值。频率值"远离"概率值的可能性永远存在，但这种可能性确实是越来越小。因此，在依概率收敛的意义下，概率可解释为频率，

反映了常量与变量的辩证统一。

(3)“比例数”是静态的概念，“概率”是动态的概念，在古典概率计算中，体现了“动”与“静”的辩证观。

例如：一袋中装有 10 个大小形状完全相同的球，其中 1~3 号是白球，4~10 号是红球，记 A 表示从袋中任摸一球是红球的事件，则

$$P(A)=7/10.$$

在这里“静态”地讲，袋中红球所占比例数为 7/10，“动态”地讲，从袋中任取一球是红球的概率为 7/10，不难看出，在“静态”向“随机”转化时，“比例”相应于“概率”。

(4)概率论在研究随机现象的方法策略也上体现了静与动的辩证统一。随机事件是从静态的观点研究随机现象，随机变量是从动态的观点研究随机现象，两者的区别也像数学中的常量与变量。构造随机变量这个抽象模型，实现了随机现象的数量化处理，使其研究方法由静态转为动态，“静”与“动”相互关联，共同揭示随机现象的统计规律。使概率论从计算孤立事件的概率走向了更高的理论体系。

(5)二项分布属于离散型，正态分布属于连续型，棣莫弗-拉普拉斯中心极限定理表述了二项分布的极限分布是正态分布，体现了离散与连续的辩证统一关系。

(6)统计推断的思想方法同样有着深刻的哲理，从局部推断整体，可以说是数理统计的一个特点。它是在对有关信息缺乏完全掌握的情况下进行的归纳推理方法，是一种定量的推理技术，从局部观察要对总体下结论，这种推理的可行性与可靠性，尚有赖于局部样本（一个个体）个性（特殊性）和总体共性（普遍性）之间的一种内在的对立统一的辩证关系。如果抽样和推理完全建立在可靠的科学基础之上（即按随机性原则抽样，加上科学的推理方法），则对总体的推断是可能的，而且结论是可靠的。统计推断的思想方法反映了“每一事物内部不但包括了矛盾的特殊性，而且包含矛盾的普遍性，普遍性存在于特殊性之中”。统计推断的结论，往往是在某一“可靠性水平”下给出的，这种矛盾的特殊性与普遍性的辩证统一，贯穿于数理统计的始终。

2.辩证唯物主义认识论教育功能

概率的观点，统计的观点以及统计推理原理，对于学生进行辩证唯物主义认识论教育有着重要作用。

(1)随机现象在一次试验中发生与否是事先不可预知的，具有偶然性，但随着试验次数的增大，它发生的次数却具有规律性。然而，即使我们已知一事件发生的概率，但仍不能肯定该事件在下一次试验中一定发生或一定不发生，此乃“在混沌中蕴涵着规律，在规律中又蕴涵着意外”。统计规律启示人们：看问题不可以绝对，思想上不拘执一端，“既要认识到一种事物从总的方面看有其一定的规律性，也承认存在例外的个案，两者看似矛盾，却是并行不悖，它反映了我们生活在其中的世界的多样性与复杂性”。

(2)概率很小的事件在一次试验中基本上不会发生，这是概率论中一条基本原理。统计推理就是基于这条原理，因而，统计推理实际上是一种合情推理（也叫探索推理、似然推理），不是逻辑推理。显著性假设检验的理论依据就是小概率事件原理，整个假设检验的逻辑程序就是带概率性质的反证法，其认识逻辑是：假定原假设成立，如果小概率事件在一次试验居然发生，那么，我们就以很大的把握否定原假设，但不是完全绝对地推翻原假设，其道理在于小概率事件在一次试验中基本不可能发生，并非绝对不能发生。当我们称

这个小概率(记为α)为显著性水平，那么，在给定显著性水平之下判断某一假设的正确性，从逻辑上讲，是一种含有否定意义的结论形式，一方面告诉你推断的结论，另一方面又告诉你这个结论有α的可能性是错误的，这种概率陈述方式揭示了随机现象具有不可控性。世间万事，能被绝对肯定和绝对否定的只是少数，而大部分是处于这两者之间，它告诉人们，凡事如果苛求获得绝对正确的结论，那你也许什么都得不到。

(3)从认识论的角度，再看统计推断方法，在许多演绎推理不可行的情况下只能进行实验，即使是演绎推理可用的场合，其结论的正确性仍须经过实验，即归纳验证。使用统计推断的方法体现了哲学上“实践第一”的观点和“实践是检验真理的唯一标准”这一辩证唯物论的认识论。

事物总是由质与量两个方面规定的，“统计方法只是从事物的外在数量表现上去研究问题，通过对数据的分析，揭示可能有某种规律性的东西存在，而不涉及事物的质的规定性。”研究事物的质是各门具体学科的事。统计分析的结果，只会告诉你从实验数据来看事情是怎样的，而不告诉你原因何在。这表明统计规律未必蕴涵因果关系，统计方法的作用在于通过数据分析，揭示表面关联的存在，指示专门研究努力的方向。只有在认识到统计方法的这一本性的前提下，才能有效应用统计方法，学会用统计的观点看世界。

(4)概率论与数理统计是有着强烈认识功能的思想结构，渗透在概率论与数理统计知识体系中的辩证唯物主义观点，是击败巫术、占术、占卜、神学等非科学自然观的锐利武器。

3.培养科学品质的功效及美育价值

概率论与数理统计在培养人的科学品质方面具有得天独厚的功能。世界上的两种推理，即论证推理与合情推理都表现在概率论与数理统计中，建立概率模型要用抽象思维；观察、实验、判断要用直觉思维；定理与性质的推证要用逻辑思维；解决概率问题，需要各种思维配合支持。通过体现在概率论与数理统计中的科学试验、逻辑论证、合情推理、理性批判以及数字化、模型化等科学方法的学习，有助于学生形成明辨是非、坚持真理、独立思考、树立自信的科学态度，以及良好的学习习惯和思维品质。

概率论与数理统计的形成与发展，期间经过了几代科学家的不懈努力与杰出贡献，在探求真理的道路上，他们洒下了勤奋的汗水。比如，在掷一枚均匀的硬币，求出现正面的概率时，蒲丰曾做过4040次的试验；英国科学家皮尔逊做过2组上万次的重复试验；在求圆周率π值时，蒲丰曾做了2212次投针试验；英国生物统计学家高尔顿为了发现频率具有的稳定性，他做了2000次钉板试验。科学家们这种为追求真理，不厌其烦的探索试验和忘我工作的精神与毅力，更是创造型人才必不可少的因素。在概率论与数理统计的公理、模型、典型范例以及计算方法中随处可以发现数学美的成分，是对学生进行美育的佐料。

下面列举几条：

(1)概率公理化定义是在概率的统计定义、古典定义、几何定义的基础上抽象统一而得的高级定义，其中3条公理简单、和谐、完备，是构成了概率论大厦的坚实基础，体现了数学的和谐美和统一美。

(2)概率论与数理统计中充斥着大量的随机模型（也称概率模型）。

比如，有限等可能概型、几何概型、伯努利概型、二项分布、泊松分布、均匀分布、正态分布、线性回归模型等，这些概率模型皆是对现实随机问题（情况）的简约、抽象和概括，体现了数学简单美的特征。比如“摸球模型”，我们把各种实际随机问题都可用一个袋

子和球这种简单装置的随机模型来刻划，从袋中任意地摸球，可以表现一般随机性的本身特点，袋中所装的球可以具有多种不同的颜色，用来代表不同的试验结果，而不同颜色的球的个数可以有多有少，用来反映不同结果发生的可能性大小。由此可见，一个概率模型与它概括的这一大堆随机现象相比，简单得实在令人惊讶，这就是数学简单美的魅力所在。

(3)“对称美”的计算功能。

等可能性建立在对称性的基础上，是对称性的一种后果，所谓对称的两个事件是指两者的结构完全一致，完全处于对称、平等的位置，因而两对称事件的概率自然应该相同。在问题中，通过对称性的判断，可获得一个独立求解条件，会产生简洁的思路和简便的算法。

例如：在线段 AB 上任取 3 点 X_1,X_2,X_3，求 X_2 位于 X_1,X_3 之间的概率。

如果注意到诸事件：A_1 =“X_1 位于 X_2 与 X_3 之间”，A_2=“X_2 位于 X_1 与 X_3”之间，A_3=“X_3 位于 X_1 与 X_2 之间”的对称性，即知

$$P(A_1)=P(A_2)=P(A_3),$$

再注意到诸事件的互斥性与完备性，便有

$$P(A_1)+P(A_2)+P(A_3)=1,$$

所以

$$P(A_2)=1/3.$$

回味解题思路，不难体会到对称美带给我们的智慧启迪。

(4)概率方法的奇异美。

蒲丰投针试验求 π 值是奇异美的一个典型。法国数学家蒲丰，将一根长 $2l$ 的小针投在距离为 $2a$（$a>l$）的若干等距平行线上，证明针与直线相交的概率是

$$p=\frac{2l}{\pi a}.$$

若

$$p\approx\frac{\mu}{n},$$

则得

$$\pi\approx\frac{2nl}{\mu a},$$

其中，μ 是针与该直线相交的次数。蒲丰竟然利用投针这一机会游戏的方法，获得了许多前人通过多方努力才求出来的 π 值。这一实验法，不但新颖奇妙让人叫绝，而且开创了用偶然性方法去攻克确定性计算的先河，闪耀着智慧的光辉，充分显示了数学的奇异美。

(5)概率悖论的奇异美。

概率论中，贝特朗（Bertrand）奇论：在一半径为 r 的圆中任作一弦，则弦长大于圆内接等边三角形边长 $\sqrt{3}r$ 的概率在 3 种不同的模型（样本空间）中，竟然有 3 种不同答案，分别是 1/2、1/3、1/4。这显然是概率论的悖论，它实际上是一种荒诞的美谈，在荒诞中蕴含着哲理，可以给人以启迪，也可以给人以美感。教学中如能充分挖掘这些数学美的因素，加强数学美的审美教育，必能对学生起到启迪智慧、陶冶情操、完善个性的作用。

8.2 在概率统计教学中渗透数学史的作法与体会

我从事多年的概率统计教学工作，发现学生对这门课普遍感到学习困难，特别是入门难。产生困难的原因很多，但由于对概率统计产生的历史背景和实际应用缺乏了解，对于这门课程的学习缺乏兴趣，是其中的重要原因之一。“兴趣是最好的老师”，怎样提高学生的兴趣?怎样解决学生学习这门课中的入门难问题?

1.以历史典故为红线，贯穿概率统计发生与发展的全过程

在概率论与数理统计的教学中，我在五处编撰了历史典故。它们像一条红线一样把统率统计发生和发展的历史串联起来，使学生在学习知识和方法的同时，了解概率统计发生、发展的历史脉络，得知概率统计还是一门年轻的科学，还需要不断地发展与完善，从而激发出他们学习的兴趣与热情。

首先是在概率的统计定义这一节后插入了：

历史典故 1

历史上有许多著名学者做过频率稳定性的试验。例如，德・摩根(De Morgan)，蒲丰(Buffon)，皮尔逊(Pearson)等人都做过大量的投掷硬币的试验，发现正面出现的频率稳定在0.5 左右。

大量地观察并统计婴儿的出生，发现男孩出生的频率稳定在 0.513 左右。18 世纪，法国数学家拉普拉斯对伦敦、圣彼得堡、柏林和整个法国的广大人口资料进行了研究，得出那些地区的男孩出生频率约等于 22/43。

又有人统计过某个国家无法投递的信件数，多年统计的结果发现，这类信件数在全部信件中的比例几乎保持不变，在百万分之五十左右。

接着在古典概型后插入了：

历史典故 2

17 世纪中叶，欧洲贵族们盛行掷骰子游戏。当时法国有一位贵族德•梅雷，他在掷骰子游戏时遇到了一些使他苦恼的问题，例如，他发现掷一颗骰子 4 次至少出现一次 6 点是有利的，而掷两颗骰子 24 次至少出现一次双 6 点是不利的。他解释不了这个现象的原因，于是向当时的法国数学家帕斯卡请教，帕斯卡接受了这些问题，并把它提交给另一位法国数学家费尔马互相讨论。他们频繁地通信，开始了概率论和组合论早期的研究。

在概率的公理化定义之后插入了：

历史典故 3

对某门数学学科公理化的努力，在数学的发展史上一直没有停止过。早期可以追溯到古希腊的欧几里得时代，而几何严格的公理化的工作是在 20 世纪初由希尔伯特完成的。在几何公理体系中，几个基本的元素点、直线、平面，几个基本的公理(又称公设) 组成了几何体系的基石，其它的概念和定理都由它们推导出来。

概率论在数学中算是一门年轻的学科，在 20 世纪以前，它还不是一门成熟的学科，许多基本概念还没有清楚地下定义，这种不清楚的地方常常引起奇论的产生。自然科学在上世纪初的发展对概率论提出了更高的要求，于是构造概率的公理化体系的任务摆在了数学家们的面前。前苏联科学院院士柯尔莫戈洛夫在前人研究的基础上，于 1933 年完成了

概率公理化的任务。概率论也才被公认为一门严格的数学学科。

在公理体系中，我们看到了数学的严密、数学的完美！

在连续型随机变量后插入：

历史典故 4

连续型随机变量中最常用的是正态分布的随机变量。正态分布最早是由移居英国的法国数学家棣莫佛提出的。他在 1718 年出版了《机遇论》一书，书中除了阐明复合事件的概率的计算方法之外，还提出了“正态分布”的概念，建立了正态曲线方程，并证明了 $p=1/2$ 时的二项分布的极限是正态分布，这个结果被法国数学家拉普拉斯推广到 $0<p<1$ 的所有场合。1812 年拉普拉斯的巨著《分析概率论》出版，该书全面地总结了到那时为止的概率论和数理统计方面的研究成果，并予以严密、系统的阐述和表达。在这一时期中，德国数学家高斯正式奠定了最小二乘法和误差理论的基础，他详细地研究了正态分布在误差理论中的作用。

在数理统计的初步知识后插入：

历史典故 5

数理统计是一门较年轻的数学学科，它虽然随着概率论的产生与应用在逐渐兴起，但主要的发展是从 20 世纪开始的。在早期的发展中，起领导作用的是以英国统计学家费歇尔和皮尔逊为首的统计学派。特别是费歇尔在数理统计的发展史中起到了独特的作用，他对样本的分布给出了严格的确定，对试验设计作了十分有益的研究，目前许多常用的统计方法都以他的名字命名，也有人称他为数理统计之父。其他的一些著名学者如戈塞特、沃尔德、奈曼、E.皮尔逊(K.Pearson 的儿子)以及我国的许宝騄教授，都作出了根本性的贡献。

这五处历史典故，概括地描述了概率统计发生、发展的全过程，而且是在讲述相关的知识时自然地渗透、灌输给学生，这比单纯地讲概率统计的发展史效果要好得多。

2.以人物简介为花絮，点缀描述概率统计曲折而艰辛的历史

为了介绍数学家、统计学家们在概率统计发展中所作的贡献，弘扬他们艰苦卓绝的奋斗精神，激励学生们的学习热情，在概率论与数理统计的教学中，插入了四处人物简介。

在伯努利概型一节之后插入了：

人物简介 1

雅各布·伯努利，17 世纪瑞士著名数学家。年轻时根据父亲的意愿学习神学，曾获巴塞尔大学文学硕士和神学硕士学位，同时怀着浓厚的兴趣研习数学和天文学。1687 年起任巴塞尔大学教授，在多方面作出重要贡献。对概率论也有深入研究，建立了描述独立试验序列的“伯努利概型”，提出并证明了“伯努利大数定律”。

值得一提的是，伯努利家族是一个数学家辈出的家族。除了雅各布·伯努利外，比较著名的还有约翰·伯努利，丹尼尔·伯努利。丹尼尔·伯努利在概率论中引入正态分布误差理论，发表了第一个正态分布表。

在切比雪夫不等式之后插入了：

人物简介 2

切比雪夫——俄国数学家、机械学家、教育学家。莫斯科大学毕业，1849 年获圣彼得堡大学博士学位。长期担任圣彼得堡大学教授，1853 年当选为圣彼得堡科学院院士。在数论、概率论、机械论方面有重要贡献，是圣彼得堡学派奠基人之一。他证明了伯特兰(Pertrand)

公式、关于自然数中素数分布的定理、概率论中的切比雪夫不等式、大数定律及中心极限定理。他从研究机械原理出发，利用多项式来逼近连续函数，创立了“函数逼近论”这一新的数学分支。他先后发表论文七十余篇，主要数学著作有《论素数》、《几何作图》等。

他在大学执教三十五年，功勋卓著，著作等身，高徒辈出，桃李满天下。他身有残疾，但矢志不渝，为科学、教育事业努力奋斗；他终生未娶，为科学、教育事业洒尽了全部心血。

在大数定律之后插入：

人物简介 3

辛钦，前苏联数学家、教育家。1916 毕业于莫斯科大学，1935 年获物理—数学博士学位。曾任莫斯科大学教授、数学力学研究所所长。1939 年当选苏联科学院通讯院士，1944 年当选俄罗斯联邦教育科学院院士。苏联概率论学派的代表人物之一。1933 年创立平稳随机过程理论，并取得极限定理等重要成果。对函数论、刁番都逼近论、连分数度量论等也有贡献。致力于教育工作，对改进前苏联的数学教育作出显著成绩。著有《概率论的极限理论》、《数学分析简明教程》、《连分数》等。曾获列宁勋章和劳动红旗勋章。著名的辛钦大数定律属于他的第一本著作《概率论的极限理论》中的内容。

在中心极限定理之后插入：

人物简介 4

棣莫佛，原籍法国的英国数学家，英国皇家学会会员，柏林科学院和法兰西科学院院士。主要贡献在概率论和代数方面。开创以正态分布为极限的中心极限定理的研究，提出关于复数 n 次乘方或开方的所谓棣莫佛公式，著有专著《机遇论》等。

拉普拉斯，法国数学家、天文学家、物理学家。曾就学于卡昂大学，后任巴黎军事学校教授。1773 年被选为法兰西科学院院士。1812 年出版《分析概率论》，总结了前人在概率论方面的工作，证明了重要的极限定理，发展了误差理论，引进了概率加法与乘法定理以及生成函数、数学期望等概念，还引进了现被广泛应用的“拉普拉斯变换”。

3.几点体会

通过多年来概率统计课的教学实践，对于在数学课程中渗透数学史的作法有以下的一些体会。

(1)结合课程，以史为线。

在讲数学课的同时，介绍一些数学史是非常必要的，这既可以增加学生的知识面，扩大学生的视野，还可以从这些史实中，了解相关的数学知识与方法产生的历史背景，体会其中的思想、方法和创立一门新学科的艰辛。要结合课程，以史为线，不必专门为史而史，也不必为重现历史，而重蹈历史上的曲折。一门课学完之后，学生既要有该门课的知识的线条，又要有这门学科的历史的线条。两条线既可平行也可相交、重叠。

(2)史不宜繁，点到为止。

由于学科课时数的限制，不可能用过多的时间去讲相关的数学史，所以在结合课程渗透数学史的时候，史料不宜过多过繁，应该简明扼要，点到为止。有时利用课堂时间讲，有时也可布置学生自学。

(3)结合实际，以史为鉴。

概率统计是一门应用十分广泛的学科，与日常生活、科学研究、工农业生产都有着紧

密的联系。但是概率论的发生却与赌博相关，并不十分光彩。我们要正确地引导学生联系实际，不能把概率当成赌博中的胜算术，应更多地看到在概率论基础上发展起来的数理统计在实际中的应用。由此引导学生注意到，一些开始看似纯理论的或纯游戏的知识，一旦找到实际的应用，其前景是十分光明的。

(4)学校教育，以人为本。

我多年从事教育，深知教育的重要任务是育人，培养社会主义的合格建设者和可靠接班人。这当中数学史的教育、数学家奋斗史的教育对学生尤其重要。在概率统计的教学中，我对每一届学生都要讲切比雪夫的事迹："他在大学执教三十五年，功勋卓著，著作等身，高徒辈出，桃李满天下；他身有残疾，但矢志不渝，为科学、教育事业努力奋斗；他终生未娶，为科学、教育事业洒尽了全部心血……"。也给学生讲 Student 分布(即 t 分布)的发现者戈塞特的事迹。讲费歇尔、K. 皮尔逊对数理统计的贡献，也讲中国的许宝騄教授、王梓坤、陈希孺教授所做的工作。这些历史人物的介绍，为学生树立了榜样，在学习数学知识的同时，又学会了如何做人。

附录 1 概率论发展大事记

1564 年，卡尔达诺完成《游戏机遇的学说》（The Book of Games of Chance）一书，是概率论的先声，可是直到他去世很久的 1663 年此书才发表。

1654 年，法国有个叫德•梅雷的赌徒向帕斯卡提出分赌注的问题（点数问题），1654 年 7 月 29 日，帕斯卡首先向费马写信转达了这些问题，标志概率论的诞生——帕斯卡与费马是概率论的创始人，把 1654 年 7 月 29 日作为概率论的诞生日，这是概率论发展史上具有“里程碑”式的事件。

1657 年，惠更斯的《论赌博中的计算》(On Reckoning at Games of Chance)出版，是概率论早期名著之一。在雅各布 · 伯努利的《猜度术》问世之前，一直无人能超过惠更斯的概率理论。

1705 年，雅各布 · 伯努利去世时，其手稿《猜度术》(The Art of Conjecturing)尚没整理定稿。最后在莱布尼兹的敦促下，雅各布的儿子才于 1712 年 10 月开始整理并印刷《猜度术》。 1713 年 5 月，当尼古拉斯回到巴塞尔时，《猜度术》的整理和印刷工作已接近尾声，为了使该书尽早付梓，尼古拉斯只匆匆写了一篇两页的序言，并为较严重的印刷错误编了一张勘误表。1713 年 8 月，在雅各布死后 8 年，才得以出版。这是概率论发展史上的具有“里程碑”式的著作，创立了伯努利大数定律。

1718 年，棣莫弗把《抽签的计算》修改为《机遇论》（Doctrine of Chances）出版，于 1738、1756 年出版了第 2、3 版，这也是概率论发展史上具有“里程碑”式的著作，给出独立事件的乘法定理和二项分布。

1733 年棣莫弗，1809 年高斯各自独立引进了正态分布。

1777 年，蒲丰发表了《或然性算术试验》，首先提出并且解决了现在称为著名的“蒲丰投针问题”，引入了几何概率，开始了几何概率的早期研究。

1763 年，贝叶斯的朋友发表了他的论文《论机会学说问题的求解》，现称为贝叶斯公式，尝试建立统计推断理论的基础。

1812 年，拉普拉斯出版《分析概率论》一书，这是概率论发展史上的具有“里程碑”式的著作。1814 年出版了第 2 版，其中增加了著名的绪论《概率的哲学导论》，1820 年出版了第 3 版，1886 年，该书收在了《拉普拉斯全集》的第七卷中。本书给出古典概率的定义，集古典概率论之大成，将分析工具引入概率论，为概率论的近代发展开辟了道路。

1837 年，泊松陈述了泊松大数定律。

1866 年，圣彼得堡学派创始人切比雪夫出版《论均值》，用他所创立的切比雪夫不等式建立了独立随机变量序列的大数定律。1867 年，又建立了有关各阶绝对矩一致有界的独立随机变量序列的中心极限定理。1898 年，马尔可夫做了补证。

1880 年，英国传教士傅兰雅与中国数学家华蘅芳合作翻译了第一部在中国传播的概率论著作《决疑数学》。

1889 年，贝特朗在他的《概率论》一书中，提出了现在称为“贝特朗悖论”这一问题，促使概率论公理化。

1906 年，俄国马尔可夫提出马尔可夫链概念。

1917 年，伯恩斯坦发表了论文《论概率论的公理化基础》，提出最早的概率论公理化的问题。1927 年，他的《概率论》第 1 版出版。

1929 年，莫斯科学派柯尔莫戈洛夫发表《测度的一般理论和概率论》，在测度论基础上建立概率的公理体系。

1931 年，柯尔莫戈洛夫的《概率论的解析方法》出版，阐述无后效随机过程理论。

1933 年，柯尔莫戈洛夫的《概率论基础》出版，建立了概率论的严格公理体系，部分解决了希尔伯特第 6 问题，为概率论的现代发展开辟了道路，这是概率论发展史上的具有“里程碑”式的著作。

1933 年，辛欣《概率论的极限定理》出版，创建平稳过程理论。

1938 年，莱维开始独立创立研究随机过程的新方法，1948 年出版《随机过程与布朗运动》。

1939 年，维尔引进“鞅”名称。1950 年开始，美国概率论学派的代表人物杜布对鞅概念进行了系统的研究而使鞅论成为一门独立的分支。

1942 年开始，日本数学家伊藤清引进了随机积分与随机微分方程，为新的数学分支——随机分析奠定了基础。

附录2 历史名题

在概率论的发展史上，有很多有趣的名题，列在下面，以飨读者。

公正分赌金问题

两个赌徒 A 、B 事先约定进行若干局的公平赌博(公平赌博意味着双方获胜的机会都是1/2，双方所出的赌资也都一样)，直到其中一人赢了某一事先规定好了的局数为止，比方说 s 局。现在由于某些偶然因素导致赌博无法进行，此时 A 赢了 s_1 局，B 赢了 s_2 局($s_1<s$，$s_2<s$)，问应如何分配赌本才算公平?赌本分配问题也称“点数问题”(problem of point)。

生日问题

某班有 n 个人($n \le 365$)，那么至少有两个人的生日在同一天的概率是多少。

分房问题

有 n 个人，每个人都有可能地被分配到 N 个房间中的任何一间去住($n \le N$)，那么求下列事件的概率:

(1)指定的 n 个房间各有一个人住。

(2)恰好有 n 个房间，其中各住一人。

装错信封问题

某人写了 n 封信，将其分别装入 n ($n \ge 2$)个信封，并在每个信封上分别随机地写上 n 个收信人的地址(不重复)，试求没有一个信封上所写地址正确的概率。

玻利亚罐子问题

罐子中有 a 个白球和 b 个黑球，每次从中随机抽取一个球，并连同 c 个同色球放回罐中，如此反复进行。试求在前 $n=n_1+n_2$ 次取球中，取出了 n_1 个白球和 n_2 个黑球的概率。

约会问题

两人相约0点到1点在某地会面，先到者等候令一个人10分钟，过时就离去。假设两人等可能在0点到1点任一时刻到达，求两人能会面的概率。

蒲丰投针问题

设在平面上有一组平行线，其间距都等于 a，把一根长 l($l<a$)的针随机投上去，试求这根针和一条直线相交的概率是多少。

贝特朗悖论

在半径为1的园内随机地取一条弦，问其长超过该园内接等边三角形的边长 $\sqrt{3}$ 的概率为多少?

赌徒输光问题

两个赌徒甲、乙进行一系列赌博。在每一局中甲获胜的概率为 p，乙获胜的概率为 q，$p+q=1$，每一局后，负者要付一元给胜者。如果起始时甲有资本 a 元，乙有资本 b 元，$a+b=c$，两个赌徒直到甲输光或乙输光为止，求甲输光的概率。

湖中有多少条鱼的问题

为估计湖中鱼数N，同时自湖中捕出r条鱼，做上记号后又都放回湖中，一段时间后再自湖中捕出S条鱼，结果发现有x条鱼标有记号，试根据此信息估计N的值。

附录3概率论与数理统计词汇英汉对照表

A

absolute value 绝对值
accept 接受
acceptable region 接受域
additivity 可加性
ajusted 调整的
alternative hypothesis 对立假设
analysis 分析
analysis of covariance 协方差分析
analysis of variance 方差分析
Antithetic variables 对立变量
arithmetic mean 算术平均值
association 相关性
assumption 假设
assumption checking 假设检验
availability 有效度
average 均值

B

balanced 平衡的
band 带宽
bar chart 条形图
beta-distribution 贝塔分布
between groups 组间的
bias 偏倚
binomial distribution 二项分布
binomial test 二项检验

C

calculate 计算
case 个案
category 类别
center of gravity 重心
central tendency 中心趋势
chi-square distribution 卡方分布
chi-square test 卡方检验
classify 分类
cluster analysis 聚类分析
coefficient 系数
coefficient of correlation 相关系数
collinearity 共线性
column 列
compare 比较
comparison 对照
components 构成，分量
compound 复合的
confidence interval 置信区间
consistency 一致性
constant 常数
continuous variable 连续变量
control charts 控制图
correlation 相关
covariance 协方差
covariance matrix 协方差矩阵
critical point 临界点
critical value 临界值
crosstab 列联表
cubic 三次的，立方的
cubic term 三次项
cumulative distribution function 累加分布函数
curve estimation 曲线估计

D

data 数据
default 默认的
definition 定义
deleted residual 剔除残差
density function 密度函数

dependent variable 因变量
description 描述
design of experiment 试验设计
deviations 差异
df。(degree of freedom) 自由度
diagnostic 诊断
dimension 维
discrete variable 离散变量
discriminant function 判别函数
discriminatory analysis 判别分析
distance 距离
distribution 分布
D-optimal design D-优化设计

E

equal 相等
effects of interaction 交互效应
efficiency 有效性
eigenvalue 特征值
equal size 等含量
equation 方程
error 误差
estimate 估计
estimation of parameters 参数估计
estimations 估计量
evaluate 衡量
exact value 精确值
expectation 期望
expected value 期望值
exponential 指数的
exponential distributon 指数分布
extreme value 极值

F

factor 因素，因子
factor analysis 因子分析
factor score 因子得分
factorial designs 析因设计
factorial experiment 析因试验
fit 拟合
fitted line 拟合线
fitted value 拟合值
fixed model 固定模型
fixed variable 固定变量
fractional factorial design 部分析因设计
frequency 频数
F-test F 检验
full factorial design 完全析因设计
function 函数

G

gamma distribution 伽玛分布
geometric mean 几何均值
group 组

H

harmonic mean 调和均值
heterogeneity 不齐性
histogram 直方图
homogeneity 齐性
homogeneity of variance 方差齐性
hypothesis 假设
hypothesis test 假设检验

I

independence 独立
independent variable 自变量
independent-samples 独立样本
index 指数
index of correlation 相关指数
interaction 交互作用
interclass correlation 组内相关
interval estimate 区间估计
intraclass correlation 组间相关
inverse 倒数的
iterate 迭代

K

kernal 核
Kolmogorov-Smirnov test 柯尔莫哥洛夫-斯米诺夫检验
kurtosis 峰度

L

large sample problem 大样本问题

layer 层
least-significant difference 最小显著差数
least-square estimation 最小二乘估计
least-square method 最小二乘法
level 水平
level of significance 显著性水平
leverage value 中心化杠杆值
life 寿命
life test 寿命试验
likelihood function 似然函数
likelihood ratio test 似然比检验
linear 线性的
linear estimator 线性估计
linear model 线性模型
linear regression 线性回归
linear relation 线性关系
linear term 线性项
logarithmic 对数的
logarithms 对数
logistic 逻辑的
lost function 损失函数

M

main effect 主效应
matrix 矩阵
maximum 最大值
maximum likelihood estimation 极大似然估计
mean squared deviation(MSD) 均方差
mean sum of square 均方和
measure 衡量
media 中位数
M-estimator M 估计
minimum 最小值
missing values 缺失值
mixed model 混合模型
mode 众数
model 模型
Monte Carle method 蒙特卡罗法
moving average 移动平均值
multicollinearity 多元共线性
multiple comparison 多重比较
multiple correlation 多重相关
multiple correlation coefficient 复相关系数
multiple correlation coefficient 多元相关系数
multiple regression analysis 多元回归分析
multiple regression equation 多元回归方程
multiple response 多响应
multivariate analysis 多元分析

N

negative relationship 负相关
nonadditively 不可加性
nonlinear 非线性
nonlinear regression 非线性回归
noparametric tests 非参数检验
normal distribution 正态分布
null hypothesis 零假设
number of cases 个案数

O

one-sample 单样本
one-tailed test 单侧检验
one-way ANOVA 单向方差分析
one-way classification 单向分类
optimal 优化的
optimum allocation 最优配制
order 排序
order statistics 次序统计量
origin 原点
orthogonal 正交的
outliers 异常值

P

paired observations 成对观测数据
paired-sample 成对样本
parameter 参数
parameter estimation 参数估计
partial correlation 偏相关
partial correlation coefficient 偏相关系数
partial regression coefficient 偏回归系数
percent 百分数
percentiles 百分位数

pie chart 饼图
point estimate 点估计
poisson distribution 泊松分布
polynomial curve 多项式曲线
polynomial regression 多项式回归
polynomials 多项式
positive relationship 正相关
power 幂
P-P plot P-P 概率图
predict 预测
predicted value 预测值
prediction intervals 预测区间
principal component analysis 主成分分析
proability 概率
probability density function 概率密度函数
probit analysis 概率分析
proportion 比例

Q

qadratic 二次的
Q-Q plot Q-Q 概率图
quadratic term 二次项
quality control 质量控制
quantitative 数量的，度量的
quartiles 四分位数

R

random 随机的
random number 随机数
random sampling 随机取样
random seed 随机数种子
random variable 随机变量
randomization 随机化
range 极差
rank 秩
rank correlation 秩相关
rank statistic 秩统计量
regression analysis 回归分析
regression coefficient 回归系数
regression line 回归线
reject 拒绝
rejection region 拒绝域
relationship 关系
reliability 可靠性
repeated 重复的
report 报告，报表
residual 残差
residual sum of squares 剩余平方和
response 响应
risk function 风险函数
robustness 稳健性
root mean square 标准差
row 行
run 游程
run test 游程检验

S

sample 样本
sample size 样本容量
sample space 样本空间
sampling 取样
sampling inspection 抽样检验
scatter chart 散点图
S-curve S 形曲线
separately 单独地
sets 集合
sign test 符号检验
significance 显著性
significance level 显著性水平
significance testing 显著性检验
significant 显著的，有效的
significant digits 有效数字
skewed distribution 偏态分布
skewness 偏度
small sample problem 小样本问题
smooth 平滑
sort 排序
soruces of variation 方差来源
space 空间
spread 扩展
square 平方

standard deviation 标准离差
standard error of mean 均值的标准误差
standardization 标准化
standardize 标准化
statistic 统计量
statistical quality control 统计质量控制
std。 residual 标准残差
stepwise regression analysis 逐步回归
stimulus 刺激
strong assumption 强假设
stud。 deleted residual 学生化剔除残差
stud。 residual 学生化残差
subsamples 次级样本
sufficient statistic 充分统计量
sum 和
sum of squares 平方和
summary 概括，综述

T

table 表
t-distribution t 分布
test 检验
test criterion 检验判据
test for linearity 线性检验
test of goodness of fit 拟合优度检验
test of homogeneity 齐性检验
test of independence 独立性检验
test rules 检验法则
test statistics 检验统计量
testing function 检验函数
time series 时间序列
tolerance limits 容许限
total 总共，和
transformation 转换
treatment 处理
trimmed mean 截尾均值
true value 真值
t-test t 检验
two-tailed test 双侧检验

U

unbalanced 不平衡的
unbiased estimation 无偏估计
unbiasedness 无偏性
uniform distribution 均匀分布

V

value of estimator 估计值
variable 变量
variance 方差
variance components 方差分量
variance ratio 方差比
various 不同的
vector 向量

W

weight 加权，权重
weighted average 加权平均值
within groups 组内的

Z

Z score Z 分数
Zeta distribution ξ分布
Zipf distribution ξ分布

参考文献

[1]A.Hald.A History of Probability and Statistics and Their Applications before 1750. New York:John Wiley and Sons Inc，1990.
[2]A.亚历山大洛夫等著，孙小礼等译. 数学——它的内容、方法和意义(第 2 卷). 北京:科学出版社，1963 年.
[3]B. B.鲍尔加尔斯基著，潘德松译. 数学简史. 上海: 知识出版社，1984.
[4]E.S.Person and M.G.Kendall.Studies in the history of statistics and probability. Charles Griffin & Company Limited London & High Wycombe，1970.
[5]F.N.David.Gods and Gambling.New York:Hafner Publishing Company，1962.
[6]G.Cardano.The Book on Games of Chance.New York:Holt，Rinehart and Winston，1961.
[7]Howard Eves 著，欧阳绛等译. 数学史上的里程碑. 北京:北京科学技术出版社，1990.
[8]H.Cramer 著，魏宗舒译. 统计学数学方法. 上海:上海科学技术出版社，1966.
[9]I.Todhunter.A History of the Mathematical of Theory of Probability from the Times of Pascal to That of Laplace.NewYork:Chelsea，1965.
[10]Ian Hacking.The Emergence of Probability.London:Cambridge University Press，1975.
[11]Ian Hacking.Abraham De Moivre.Discovery of Scientific Biograph，1974(19):135~136.
[12]Ian Hacking.An Introduction to Probability and Inductive Logic. London:Cambridge University Press，2001.
[13]Ian Hacking.Jacques Bernoulli' s Art of Conjecturing.British J.Philos.Sci.，1971(22):209~229.
[14]J.Bernoulli.Ars Conjectandi.Basel，1713.
[15]J.Leroy Forks，魏宗舒等译. 统计思想. 上海:上海翻译出版社，1987.
[16]John Tabak 著，杨静译. 概率论和统计学——不确定性的科学. 北京:商务印书馆，2007.
[17]Karl Person.The History of Statistics in the 17th and 18th Centuries，Against the Changing Bachground of Intellectual，Scientific and Religious Thought.Charles Griffin & Company Limited London & High Wycombe，1978.
[18]L.Daston.Classical Probability in the Enlightenment，Princeton. Princeton University Press，1988.
[19]M.Kline 著，齐民友等译. 现代世界中的数学. 上海:上海教育出版社，2004.
[20]M.Kline 著，张理京等译. 古今数学思想(第 1~4 册) . 上海：上海科学技术出版社，1979.
[21]M.Kline 著，李宏魁译. 数学:确定性的丧失. 长沙:湖南科学技术出版社，2004.
[22]Olav Kallenberg.Foundations of Modern Probability(影印版). 北京:科学出版社，2001.
[23]P.S.Laplace.Philosophical Essay on Probability. New York: Springer-Verlag，1995.
[24]P.S.Laplace.Theorie Analytique Probabilities. Paris: Courcier，1812.
[25]R.A.Fisher(1912).On an Absolute Criterion for Fitting Frequency Curves.Messenger of Mathematics41，155~160.
[26]R.A.Fisher(1922).On the Mathematical Foundation of Theoretical Statistics.Philos.Trans.Roy.Soc.London Ser.A222，309~368.
[27]R.A.Fisher(1925).The Goodness of Fit of Regression Formulae，and the Distribution of Regression Coefficients.J.R.Stat.Soc.85，597~612.

[28]R.A.Fisher(1925).Theory of Statistical Estimation.Proc.Cambridge Philos.Soc.22，700~725.
[29]R.A.Fisher(1925).Statistical Methods for Research Workers.Olivers & Boyd，Edinburgh.
[30]Sheldon Ross 著，赵选民等译. 概率论基础教程. 北京:机械工业出版社， 2006.
[31]Victor J.Katz 著，李文林等译. 数学史通论(第 2 版). 北京:高等教育出版社，2004.
[32]William Feller 著，胡迪鹤译. 概率论及其应用（第 3 版）. 北京:人民邮电出版社，2006.
[33]Zhu Chunhao.Asymptotic Normality of Multi-dimension Quasi Maximun Likelihood Estimate in Generalized Linear Models with Adaptive Design.高等数学通报，2006（4）:58~63.
[34]陈希孺. 数理统计学简史. 湖南:湖南教育出版社，2002.
[35]陈希孺. 数理统计学小史（一）. 数理统计与管理，1998(2):60~65.
[36]陈希孺. 数理统计学小史(二). 数理统计与管理，1998(3):61~64.
[37]陈希孺. 数理统计学小史(三). 数理统计与管理，1998(4):54~61.
[38]陈希孺. 数理统计学小史(最小二乘法). 数理统计与管理，1998(6):56~62.
[39]陈希孺. 最小一乘线性回归(上). 数理统计与管理，1989(5):48~55.
[40]陈希孺. 最小一乘线性回归（下）. 数理统计与管理，1989(6):48~56.
[41]陈希孺. 数理统计学:世纪末的回顾与展望. 统计研究，2000(2).
[42]陈景寅. 论概率统计与美. 济南大学学报，1996(4):73~76.
[43]蔡东汉. 概率论简史. 华中师范大学学报(自然科学版)，1990(4):516~519.
[44]方开泰，吴传义等. 数理统计与标准化. 北京:技术标准出版社，1981.
[45]丰璐. 概率论的缘起和发展. 中学数学杂志(高中)，2005(1):65.
[46]葛余博. 赵衡秀. 概率论与数理统计. 北京:科学出版社，2003.
[47]葛余博. 概率论与数理统计. 北京:清华大学出版社，2005.
[48]龚鉴尧. 世界统计名人传记. 北京:中国统计出版社，2000.
[49]高庆丰. 欧美统计学史. 北京:中国统计出版社，1987.
[50]高隆昌. 数学及其认识. 北京:高等教育出版社，施普林格出版社，2001.
[51]郭贵春，宋尚玮. 对概率论起源的思考. 科学技术与辩证法，2006(2):89~93.
[52]郭世荣. 西方传入我国的第一部概率论专著——《决疑数学》. 中国科技史料，1989(2):90~96.
[53]郭照庄，孙月芳，张良勇. 浅谈概率的对象及其意义. 高等函授学报(自然科学版)，2006(5):54~56.
[54]黄萍. 解读赖欣巴哈的逻辑哲学思想——评述其概率归纳逻辑理论及概率的意义论. 西南大学硕士学位论文，2006.
[55]黄晶晶，黄世同. 关于贝特朗悖论的新思考. 昆明师范高等专科学校学报，2004(4):10~12.
[56]黄书亭，刘波. 数学分析对概率论的渗透与推动. 自然辩证法研究，1995(4): 15~21.
[57]黄书亭. 概率四论. 工科数学，1991(4):132~136.
[58]黄汉平. 现代概率论的奠基人——柯尔莫戈洛夫. 自然杂志，1990(3): 177~179.
[59]何晓峰. 惠更斯与经典概率论. 自然杂志，1989(8):623~626.
[60]贾小勇，徐传胜，白欣. 最小二乘法的创立及其思想方法. 西北大学学报(自然科学版)，2006(3):507~511.
[61]江金彦. 随机性的本质是什么. 统计与信息论坛，2002(3):29~30.
[62]解延年. 切比雪夫. 数学通报，1987(1):42~43.

[63]柳延延．现代科学方法的两个源头——统计学与概率论的产生与结合．自然科学史研究，1996(4):309~318.
[64]柳延延．决定论之谜——复杂性科学中的偶然性与必然性．自然辩证法研究，1990(2):18~24.
[65]李永国．各类分布产生的背景．东北师范大学硕士学位论文，2006.
[66]李文林．数学珍宝．北京:科学出版社，1998.
[67]李文林．数学史教程．北京:高等教育出版社，施普林格出版社，2000.
[68]李文林．数学的进化——东西方数学史比较研究．北京:科学出版社，2005.
[69]李俊德．概率论和统计学的起源、发展、形成和现状．数学通报，1990(9): 46~47.
[70]李斌．浅谈随机思想在概率论教学中的作用．宿州学院学报，2006(4): 119~121.
[71]李兆兴．概率论中三个重要分布相互关系的一点讨论．大庆师专学报，1993(4):18~20.
[72]李敬革，王玉梅．拉普拉斯决定论的成因及其历史地位．自然辩证法研究，1994(9):24~30.
[73]李明．概率论与数理统计教学方法之体会．统计教育，2006(7):22~23.
[74]刘森．概率论与数学分析知识的相互运用．伊犁师范学院学报，2006(3): 5~9.
[75]李惠村．欧美统计学派发展简史．北京:中国统计出版社，1984.
[76]蔺云．哲学与文化视角下概率统计课的育人功能．数学教育学报，2002(2): 24~26.
[77]蔺云，巫伟科．转换与变换是概率论发展的杠杆．嘉应大学学报(自然科学版)，2003(6):14~18.
[78]赖景耀．概率论的起源与发展．西北师范学院学报，1984(3):10~16.
[79]赖景耀，王春林．概率论导引．兰州:甘肃教育出版社，1996.
[80]雷旭辉，许涤龙．概率论在统计学科发展中的角色研究．统计教育，2005(5): 1~3.
[81]茆诗松，王静龙，濮晓龙．高等数理统计(第 2 版)．北京:人民教育出版社， 1989.
[82]乔文进．论数学在法律中的应用．苏州大学硕士学位论文，2004.
[83]钱辉萍．浅谈几何概率．高中数学教与学，2005(12):40~41.
[84]舒爱莲．古典概率思想的发展过程及要义．山东轻工业学院学报，2006(2): 77~81.
[85]舒爱莲．论西方数学家偶然性中求必然、不确定中求确定的早期探索．山东大学硕士学位论文，2005.
[86]盛骤等．概率论与数理统计(第 3 版)．北京:高等教育出版社，2003.
[87]宋尚玮．对概率论起源的哲学思考．山西大学硕士学位论文，2006.
[88]宋作祥．囚徒困境与数学期望．高中数学教与学，2005(8):11~12.
[89]苏淳．概率论．北京:科学出版社，2004.
[90]孙荣恒．趣味随机问题．北京:科学出版社，2004.
[91]孙雨仁．哲学在概率论学科发展中的作用．贵阳师专学报(社会科学版)，1996(2):69~73.
[92]孙道德．从概率论的发展与应用谈其实质性．阜阳师范学院学报(自然科学版)，1994(1):24~30.
[93]孙淑珍．概率论发展轨迹的启示．郑州轻工业学院学报，1987(2):57~63.
[94]史及民．关于 Buffon 掷针问题的深思．山西师大学报(自然科学版)， 1996(2):1~3.
[95]石莹．概率论・对称思想・中学数学．数学教育学报，1998(3):85~87.
[96]吴文俊．世界著名数学家传记．北京:科学出版社，1990.
[97]吴文俊．中国数学史大系(第八卷)．北京:北京师范大学出版社，2000.

[98]吴彤，于金龙．柯尔莫戈洛夫:复杂性研究的逻辑建构过程述评．自然辩证法研究，2003(9):71~74.
[99]魏宗舒．概率论与数理统计教程．北京:高等教育出版社，1980.
[100]王梓坤．概率论及其应用．北京:科学出版社，1972.
[101]王树禾．数学思想史．北京:国防工业出版社，2003.
[102]王寿仁．概率论概貌．北京:科学技术文献出版社，1991.
[103]王幼军．拉普拉斯概率理论的历史研究．上海交通大学博士学位论文，2003.
[104]王幼军．审视概率革命．自然辩证法研究，2002(11):29~32.
[105]王幼军．概率论的起源——机会性游戏．数学教学，2004(6):44~47.
[106]王幼军．《决疑数学》——一部拉普拉斯概率论风格的著作．自然科学史研究，2006(2):159~169.
[107]王幼军．拉普拉斯概率论的衰落．自然辩证法研究，2007(3):100~103.
[108]王丽霞，杨静．人类对于随机性认识的四个阶段．自然辩证法通讯，2006(3):62~66.
[109]王兰措．概率论的发展及应用．青海民族学院学报，1996(3):110~111.
[110]王晓勤，韩祥临．中学数学中的数学史．北京:科学出版社，2002.
[111]王青建．卡尔达诺的概率著作初探．数学史研究文集(第六辑)，1998:96~101.
[112]《现代应用数学手册》编委会．现代应用数学手册(概率统计与随机过程卷)．北京:清华大学出版社，2000.
[113]肖云茹．概率统计计算．天津:南开大学出版社，2001.
[114]谢文耀．华衡芳和《决疑数学》．中国统计，1991(10):44~45.
[115]学哲学小组．对概率论发展的几个问题的认识．数学学报，1976(2):73~82.
[116]徐伯华．概率论诞生记．中学数学研究，2006(3):47~49.
[117]徐伯华．概率论诞生的思想历程．咸阳师范学院学报，2006(4):16~19.
[118]徐洪香．概率论的缘起、发展及其应用．辽宁工学院学报，2001(3):62~63.
[119]徐传胜，曲安京．惠更斯与概率论的奠基．自然辩证法通讯，2006(6):76~80.
[120]徐传胜，曲安京．许宝騄对概率论与数理统计的卓越贡献．中国科技史杂志，2006(4):340~347.
[121]徐传胜．概率论与红楼梦．数学通报，2004(1):36~38.
[122]徐传胜．概率论简史．数学通报，2004(10):36~39.
[123]徐传胜．康熙大帝与数学大师．数学通讯，2003(Z1):25~26.
[124]徐传胜．切比雪夫的概率思想及其数学文化背景．自然辩证法研究，2005(7): 29~33.
[125]徐传胜，傅尊伟．概率问题中的某些积分技巧．高等数学研究，2004(4): 31~33.
[126]徐传胜．离散型随机变量数学期望的求法探究．高等数学研究，2005(1): 33~35.
[127]徐传胜．运用实际问题改进《概率统计》教学．数学教育学报，2000(4): 91~94.
[128]徐传胜，曲安京．拉普拉斯的《分析概率论》研究．自然科学史研究，2006(3):227~238.
[129]徐传胜，吕建荣．亚伯拉罕・棣莫弗的概率思想与正态概率曲线．西北大学学报(自然科学版)，2006(2):339~343.
[130]徐传胜．雅各布•伯努利的《猜度术》研究．数学研究与评论，2007(1): 212~218.
[131]徐传胜，郭政．数理统计学的发展历程．高等数学研究，2007(1):121~125.
[132]徐传胜，潘丽云，任瑞芳．惠更斯的 14 个概率问题研究．西北大学学报(自然科学版)，2007(1):170~174.

[133]徐传胜，张梅东．正态分布两发现过程的数学文化比较．纯粹数学与应用数学，2007(1):137~144.
[134]严士健，王隽骧，刘秀芳．概率论基础．北京:科学出版社，1982.
[135]余锦华等．概率论与数理统计．广州:中山大学出版社.
[136]于忠义．人类早期不确定性推断思想研究．天津财经学院博士学位论文，2003.
[137]于忠义．高斯与观测误差分布的发现．统计与信息论坛，2006(6):28~30.
[138]于忠义．James Bernoulli 与《推测术》．统计研究，2003(5):59~61.
[139]殷烁，丛玉华，何淑华．主观概率进入概率论教材的必要性．通化师范学院学报，2005(6):101~102.
[140]杨静．概率论思想的历史演变．河北师范大学硕士学位论文，2003.
[141]杨向群等．德识才学的实践者——庆贺王梓坤院士 75 岁华诞．应用概率统计，2004(3):334~336.
[142]袁卫．统计推断思想．北京:中国统计出版社，1990.
[143]钟开莱，魏宗舒等译．初等概率论附随机过程．北京:人民教育出版社，1980.
[144]钟开莱．许宝騄在概率论方面的工作．数学的实践与认识，1980(3):12~15.
[145]张德然．概率论思维论．合肥:中国科学技术大学出版社，2006.
[146]张栋栋，张德然．概率论思维及其智力品质的培养．大学数学，2005(5): 103~108.
[147]张顺燕．数学的思想、方法和应用．北京:北京大学出版社，1997.
[148]张驰．概率论导引．成都:四川大学出版社，2001.
[149]张驰．在概率统计教学中渗透数学史的作法与体会．高等教育研究，2006(1):57~59.
[150]张炎．“实践——认识——实践”是概率论历史发展的必然规律．四川师范学院学报(高教研究专号)，1994(6):65~67.
[151]张奠宙．20 世纪数学经纬．上海:华东师范大学出版社，2002.
[152]张奠宙．中国近现代数学的发展．石家庄:河北科学技术出版社，2000.
[153]朱家生．数学史．北京:高等教育出版社，2004.
[154]朱冬梅．谈概率论中三种重要的分布．开封教育学院学报，2003(4):42~43.
[155]朱春浩. 正态变换及其误差比较研究. 山东师范大学学报(自然科学版),2006(4):11~12.
[156]朱春浩．自适应设计广义线性回归多维拟似然估计的渐近正态性．经济数学，2006(4):400~406.
[157]朱春浩．误差为鞅差序列的回归函数估计的收敛速度．经济数学，2007(1):75~81.
[158]朱春浩．鞅差序列的 Bernstein 型不等式及其应用．高等数学通报，2007(1):53~57.
[159]朱春浩．随机截断下 NA 样本半参数回归模型中的相合估计．数学杂志，2007(3):327~332.
[160]朱春浩．误差为鞅差序列的回归函数估计的强相合型．新疆师范大学学报（自然科学版），2007(2):6~12.
[161]朱春浩．最小一乘法与最小二乘法:历史与差异．统计与决策，2007(6):9~10.
[162]赵林城．学而不厌诲人不倦——祝贺陈希孺院士七十华诞．应用概率统计，2004(1):105~108.
[163]中国大百科全书总编辑委员会《数学》编辑委员会．中国大百科全书(数学)．北京:中国大百科全书出版社，1988.

［Ｇｅｎｅｒａｌ　Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝概率论思想方法的历史研究
作者＝朱春浩编著
页数＝２２７
ＳＳ号＝１１８８６５１５
出版日期＝２００７年０８月第１版